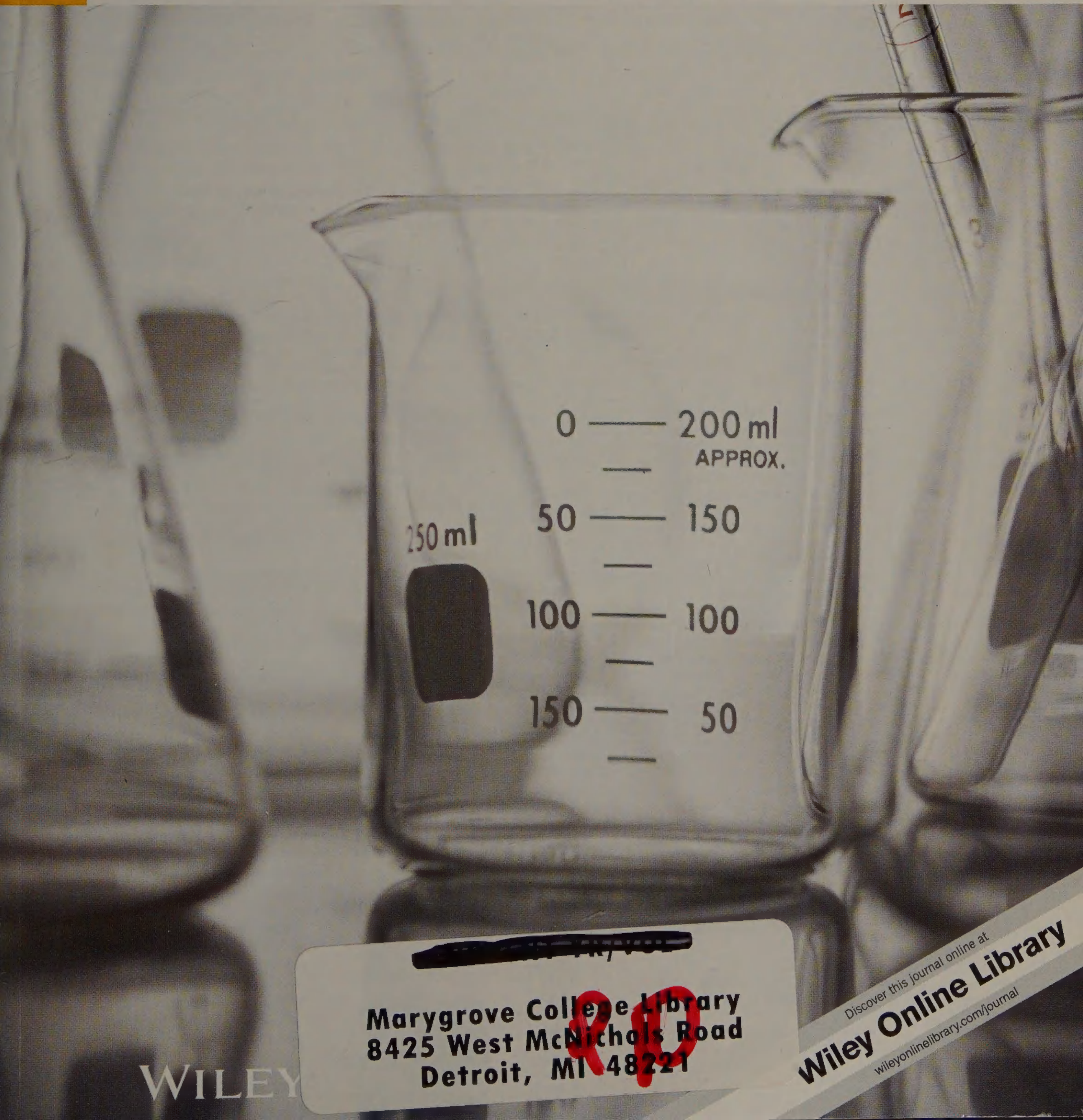
Science Education

MARCH 2014 / VOLUME 98 / ISSUE NO 2



WILEY

Science Education

SCIENCE EDUCATION (ISSN 0036-8326 (Print); ISSN 1098-237X (Online)) is published bimonthly by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Periodical Postage Paid at Hoboken, NJ and additional offices. **Postmaster:** Send all address changes to SCIENCE EDUCATION, Journal Customer Services, John Wiley & Sons Inc., 350 Main St., Malden, MA 02148-5020.

Wiley's Corporate Citizenship initiative seeks to address the environmental, social, economic, and ethical challenges faced in our business and which are important to our diverse stakeholder groups. Since launching the initiative, we have focused on sharing our content with those in need, enhancing community philanthropy, reducing our carbon impact, creating global guidelines and best practices for paper use, establishing a vendor code of ethics, and engaging our colleagues and other stakeholders in our efforts. Follow our progress at www.wiley.com/go/citizenship
Information for subscribers: SCIENCE EDUCATION is published in 6 issues per year. Institutional subscription prices for 2014 are: Print & Online: US$2107 (US), US$2233 (Rest of World), €1456 (Europe), £1152 (UK). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST/HST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to wileyonlinelibrary.com/tax-vat. The price includes online access to the current and all online back files to January 1st 2010, where available. For other pricing options, including access information and terms and conditions, please visit wileyonline library.com/access. **Delivery Terms and Legal Title:** Where the subscription price includes print issues and delivery is to the recipient's address, delivery terms are Delivered at Place (DAP); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers FOB our shipping point, freight prepaid. We will endeavour to fulfil claims for missing or damaged copies within six months of publication, within our reasonable discretion and subject to availability. **Journal Customer Services:** For ordering information, claims and any enquiry concerning your journal subscription please go to www.wileycustomerhelp.com/ask or contact your nearest office. **Americas:** Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or +1 800 835 6770 (toll free in the USA & Canada). **Europe, Middle East and Africa:** Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315. **Asia Pacific:** Email: cs-journals@wiley.com; Tel: +65 6511 8000. **Japan:** For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free) 005 316 50 480. **Visit our Online Customer Get-Help** available in 6 languages at www.wileycustomerhelp.com
Back issues: Single issues from current and recent volumes are available at the current single issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com
Advertising Sales: Inquiries concerning advertising should be forwarded to Advertising Sales Manager, Joe Tomaszewski, John Wiley & Sons, 111 River Street, Hoboken, NJ 07030: (201) 748-8895. Email: jtomaszews@wiley.com
Production Editor: Carol Neff (carol.neff@aptaracorp.com).
Other correspondence: All other correspondence should be addressed to Science Education, Publisher, c/o John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030.
Information: For submission instructions, subscription, and all other information and to view this journal online, visit: wileyonlinelibrary.com/sce
Online Open: SCIENCE EDUCATION accepts articles for Open Access publication. Please visit http://olabout.wiley.com/WileyCDA/Section/id-406241.html for further information about OnlineOpen.

This journal is indexed by *CIJE: Current Index to Journals in Education (ERIC), Contents Pages in Education (T&F), Current Abstracts (EBSCO), Current Contents® /Social & Behavioral Sciences (Thomson ISI), Education Index/Abstracts (HW Wilson), Educational Research Abstracts Online (T&F), ERIC Database (Education Resources Information Center), FRANCIS Database (INIST/CNRS), IBR & IBZ: International Bibliographies of Period-ical Literature (KG Saur), Informal Science.org (University of Pittsburgh Center for Learning in Out-of-School Environments), Journal Citation Reports/Social Science Edition (Thomson ISI), Psychological Abstracts/PsycINFO (APA), Research into Higher Education Abstracts (T&F), SCOPUS (Elsevier), Social Sciences Citation Index® (Thomson ISI), Social SciSearch® (Thomson ISI), and Web of Science® (Thomson ISI).*
Printed in the USA by The Sheridan Group.

♾ **This paper meets the requirements of ANI/NISO Z39.48-1992 (Permanence of paper).**

VOLUME 98, ISSUE NUMBER 2 MARCH 2014

Science Education

CONTENTS

Science Education

Science
Education

Children's Motivation Toward Science Across Contexts, Manner of Interaction, and Topic

MEGHAN E. BATHGATE, CHRISTIAN D. SCHUNN, RICHARD CORRENTI
Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received 20 August 2013; accepted 29 October 2013
DOI 10.1002/sce.21095
Published online 31 December 2013 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Understanding the features of science learning experiences that organize and motivate children at early ages can help educators and researchers find ways to ignite interest to support future passion and learning in the sciences at a time when children's motivation is declining. Using a sample of 252 fifth- and sixth-grade students, we systematically explore differences in children's motivations toward science experiences across context (formal, informal, neutral), manner of interaction (consuming new knowledge, analyzing, action), and topic (e.g., biology, earth science, physics). Motivations toward science were most influenced by topic. Responses were generally consistent across context and manner of interaction. Implications for science education, as well as measurement and assessment methodology, are discussed. © 2013 Wiley Periodicals, Inc. *Sci Ed* **98:**189–215, 2014

INTRODUCTION

Despite society's growth in scientific knowledge, research shows a gradual decline in children's motivation toward science as they approach adolescence (e.g., Osborne, Simon, & Collins, 2003; Simpson & Oliver, 1990; H. T. Zimmerman, 2012). Unfortunately, this decrease coincides with the sensitive timing of science choices and milestones that are influential to future science opportunities, such as science camps, advanced science courses, and calculus (Tyson, 2011; Tyson, Lee, Borman, & Hanson, 2007). As a result, a number of children will have made early experience choices that limit what they can later do based

Correspondence to: Meghan Bathgate; e-mail: meb139@pitt.edu

on premature evaluations of their fit with science (e.g., on the basis of stereotypes). Such early influences on career decision making could prevent the diversification of the pool of scientists and engineers that is currently being sought because obtaining a formal career in science depends upon the cumulative impact of all of these choices (Adams et al., 2011; Archer et al., 2012). Furthermore, reductions in openness and curiosity toward science experiences may prevent many children from fully developing scientific literacy, reducing what they can understand about technology, medical issues, and environment concerns as adults. Understanding why and in what ways children's motivation in science drops during early adolescence can help us learn how to mediate the decline in science across this age.

A number of efforts to increase individuals' science literacy and participation in science-related careers have focused on early exposure to science with the aim of generating long-term interest toward science. Hidi and Renninger (2006) suggest that general interest builds from interest in particular situations. Such a shift from particular to general seems intuitive at first blush, but actually hides a number of key complexities. Vis-à-vis a complex social construct like science, what is the character of those particular situations in the mind of the child? Science can be conceived of as a set of topics, a set of activities, and a set of places of engagement. For example, in developing a relationship to science, do children focus on the kind of tasks they are asked to do in that situation (e.g., hands- on science)? Or do they focus on the topic of inquiry (e.g., dinosaurs)? Or is the context the salient element (e.g., science camp or the class period called "science")? The ways in which children generalize early positive or negative experiences with science will likely be heavily influenced by the ways in which such situations are represented (Eshach & Fried, 2005). At the same time, the regularities in their environments will likely also shape the scope of interests and motivation that children have toward science (e.g., all the classroom-based experiences were dull or all the dinosaur experiences were exciting). We investigate how these aspects of science frame motivations in science.

Specifically, the goal of this paper is to investigate how students' early motivation varies across the dimensions through which science occurs. The literature suggests several frames for how a child's experience with science might be influenced, such as the context in which science is experienced (e.g., formal vs. informal spaces), the manner in which children interact with science materials or ideas (e.g., hands-on vs. worksheet activities), and the science content (e.g., physics vs. biology). These frames are our dimensions of interest. Each of these dimensions has been argued to be influential to children's science understanding and science motivation (Dierking, Falk, Rennie, Anderson, & Ellenbogen, 2003; Jacobs, Finken, Griffen, & Wrightm, 1998; Mantzicopoulos, Samarapungavan, & Patrick, 2009).

Context

The simple "formal versus informal" distinction has existed in modern learning research for a number of years, yet what is encompassed by this distinction can be defined in multiple ways (Dierking et al., 2003; Hofstein & Rosenfeld, 1996). Formal science contexts are frequently defined as school-based science experiences, leaving informal learning to include a diverse set of out-of-school science experiences. However, there are a number of elements typically associated with the formal/informal distinction, such as by the relationships of participating individuals (e.g., teacher-guided classroom instruction, peer discussion), structure of a program (e.g., highly structured with clear expectations, unstructured with no expectation), or whether the child self-selected to participate in an activity (e.g., compulsory vs. free-choice) (Dierking & Falk, 2003; Dierking et al., 2003; Vadeboncoeur,

2006). The focus on performance assessments may also vary across contexts. Classroom science often has an evaluative component in which children are asked to demonstrate their knowledge and are given feedback often in the form of grades. Many informal experiences are less individually evaluative, although these activities are not free from competition and achievement, as can be seen often with sports teams or camp competitions (Ntoumanis, Taylor, & Thogersen-Ntoumani, 2012). These common formal versus informal experience characteristics imply that children will often explore science with different interactions, constraints, and expectations.

For our current purposes, we conceptualized context into the following categories: "formal" science, related to in-class experiences; "informal" science, representing those activities occurring outside of the classroom in an environment more likely to allow for free-choice, such as at a home, camp, or with friends outside of school; and a "neutral" category to explore whether adding a context shaped responses, which did not explicitly specify a context and could be relevant to both formal and informal contexts (e.g., "Understanding science is helpful for solving problems"). While we recognize that the boundaries of "formal" and "informal" are somewhat artificial and that a child's overall science experience is cumulative across a range of spaces (Dierking & Falk, 2003), our decision to dichotomize across formal/informal spaces was motivated by two reasons. First, our research question examines a child's sensitivity toward a number of dimensions of science, broadly constructed, and thus some simplification of each dimension, including formal/informal is required to make the study manageable and sufficiently powered. Second, although school-aged children generally have some in and out-of-school science experience, children vary in their exposure to particular subtypes of in and out-of-school science experiences (Sha, Schunn, & Bathgate, 2013). Using too fine a slice to differentiate between formal (e.g., text book vs. hands-on scripted vs. project based) or informal experiences (e.g., clubs vs. summer camps vs. science at home) is not possible for children with little experience or restricted in type of experience. As such, we attempted to gather a variety of examples of each category to adequately represent typical forms and features of these contexts.

In addition, to further allow for differences in children's specific experiences in particular activities, many items were phrased as modals (e.g., "If I . . . "). Since a key function of motivation is to drive choices, if children have a clear motivational preference based on choice characteristics (e.g., type, location, and topic of activity), these preferences are consequential even if these preferences are based on little prior experience.

Manner of Interaction

What does it mean to "do" science? Science has both a declarative domain knowledge (i.e., the content of a discipline), and processes and strategies within a given domain (C. Zimmerman, 2000). Furthermore, "science" itself covers a number of disciplines, and each discipline within science involves its own processes, discourses, analytic techniques, and ways of interpreting phenomena. The manner in which scientific processes are enacted, the speed with which these processes are done, the specific tools and techniques used, how findings are communicated, and how feedback is received varies across specific disciplines of science. For example, an astrophysicist spends a good deal of time working with mathematical models on a computer, never interacting with the physical substances being studied, whereas a marine biologist often works within the environment being studied, interacting with the living organisms being studied. Alternatively, an evolutionary biologist carefully studies and analyzes the past, but cannot alter and manipulate their artifacts in the way a chemist can through a series of experiments. Each domain has a concrete set of contextual

expectations for knowledge and work that allows an individual to progress to expert levels within a narrow scope of the larger context of science. Yet, while the declarative and procedural knowledge differ across domains, there remain some shared foundational scientific processes.

How does appreciation of the scientific process, or of these distinctions within sciences, play out for children? Elementary-aged students often begin learning science domain knowledge via the classification and categorization of science content, but the more abstract science processes are not discussed until later in their education (Metz, 1995). This delay, whether a necessary step in developing scientific understanding or not, can affect a child's understanding of what science is and what it means to "do science" (Mantzicopoulos et al., 2009; Metz, 1995). This delayed appreciation of science processes may in part be attributable to the current education system in which curricula quickly cover a wide range of science content (Schmidt, McKnight, & Raizen, 1997) within a small amount of class time devoted to science. All of this is further complicated because teachers have limited knowledge of science processes (Fulp, 2002).

Furthermore, when children first begin to learn science, they are exposed to a variety of experiences that require a child to interact with science material in different ways that may or may not be authentic to the science being studied. The learning of science can rely on textbook reading, discussions, hand-on worksheet problems, inquiry investigations, or group work to explore science concepts across a range of science domains (Fulp, 2002). These modes of learning are substantially different from one another. Because children may have varying preferences with the way they interact with science material across these modes of learning, there may be large differences in the degree of engagement, interest, and understanding children have when interacting across these different experiences.

Although there is a common belief that hands-on activities are most engaging for children, research in this area shows conflicting results about the benefit of hands-on activities for learning outcomes and engagement, suggesting hands-on activities may not always be beneficial without structured, mindful guidance from instructors (Hofstein & Lunetta, 2004). Furthermore, the modes of learning can function together, moving between reading, experimentation, and discussions. Given the array of scientific activities in which a child participates of varying quality sometimes presented together and sometimes presented separately, it is an open question whether children show a preference for one type of science learning activity more strongly than another regardless of topic or context.

We divide manner of interacting into three categories, representing commonly discussed large divisions in subjective focus of the interaction type: *consuming new knowledge*, which involves the studying, reading, and going online for the learning of new science information (Hidi & Renninger, 2006); *analyzing*, which describes what may occur more within a child's mind and involves a child's thinking about information they have already learned (Mercier & Sperber, 2011; Vygotsky, 1978); and *action*, where a hands-on activity is specified (e.g., building things) (Wigfield, Guthrie, Tonks, & Perencevich, 2004). Although a given situation will often involve at least two if not all three of these categories, our distinctions here place emphasis on particular elements (the physical interaction or the acquisition of new information or the pondering of existing content) to understand how the subjective focus influences engagement.

Topics

Children are more likely to engage in learning experiences if they are interested and curious in the content (Hidi & Renninger, 2006). As science can be subdivided into different domains and is often taught in this partitioned way, researching "science" at the general level

may not be nuanced enough to discover differences in children's motivation toward science. Even by kindergarten, children express some differences in their motivation toward different science disciplines (Mantzicopoulos, Patrick, & Samarapungavan, 2008); however, there are large differences in exposure to diverse science topics. By the time children are out of middle school, larger preferences can be found across a range of scientific areas (ByBee & McCrae, 2011; Trumper, 2006a, 2006b).

These developmental differences raise the question: at what level are these differences found in children? How specific are their interests at this age? Topic differentiation may occur at a very narrow level (e.g., dinosaurs), and interest may be found only in instances in which this topic is found. Alternatively, interest areas may be broader (e.g., biology) or even expansive (e.g., all science). Developmental process and new science experiences may affect these interests. Initial interest in a science topic may be triggered by a particularly engaging experience that initiates interest, and this interest may become a more personalized and self-driven interest that develops over time (Hidi & Renninger, 2006). However, the breadth of that interest may change as a child develops his or her understanding toward different components of science and his or her affective-cognitive reaction to them.

In conceptualizing the grain size of learner preferences, it is important to consider what kinds of distinctions children are likely to make. At older ages, children and adults have (potentially strong) associations with science disciplines per se, such as loving biology but hating physics. At younger ages, children may not know the labels or even meaning of typical science disciplines, like chemistry or earth sciences. But they may have already developed affinity for a range of topics within disciplines (e.g., various biology topics they have already encountered) (Crowley & Jacobs, 2002). By breaking down these larger domains into specific instances, it is possible to probe for disciplinary interests while avoiding complex terms unfamiliar to the child. Thus, we examine this topic dimension by including a range of science topics within five large domains of science (astronomy, biology, earth science, engineering, physical science[1]). In addition, however, it is useful to consider children's motivation at the general level of "science" to understand how children's general science motivation may differ from topic-specific motivation. Children may have idiosyncratic associations with the general term, yet the world is often labeled using that term and thus it is important. As a note about terminology, we asked children about "science," but we will use the term "general science" in our discussion here to help distinguish analyzing motivations about a general label from analyzing motivations across many topics.

Finally, the motivation literatures have identified a large number of constructs that influence student participation and engagement in science. For the purposes of the current study—understanding the contextualization of motivations in science—we sampled a subset of the motivational constructs (e.g., interest, appreciation, identity) to serve as the basis for our item structure that (1) have been previously associated with outcomes such as learning, achievement, and future activity choices; (2) come from a range of theoretical perspectives, and (3) are not mutually redundant (Archer et al., 2012; Bryan, Glynn, & Kittleson, 2011; Jacobs et al., 1998; Lent, Brown, & Larkin, 1984). Specifically, our measures included items relating to children's self-reported appreciation toward science, curiosity and interest toward science, identity with science, persistence in science activities, personal responsibility for learning science, and expectancy value in science. Table 1 provides a concrete conceptualization for these constructs and key references. However, our current purpose is not to formally test the differences among motivational constructs, but

[1]"Physical science" is the label given in the United States for physics and chemistry topics at the middle-school level.

TABLE 1
Conceptualization of Each Motivational Subscale, Examples, and Citation Sources

Construct	Conceptualization	Item Example	Citation Source Example
Appreciation	Appreciation items inquired about children's understanding of the value and nature of science in their lives.	Understanding science is helpful for solving problems.	Schreiner and Sjoberg (2004); Weinburgh and Steele (2000)
Curiosity	Curiosity items were designed to assess children's wondering, investigating, and excitement in learning more about science-related topics. Specifically, items asked about children's seeking understanding, opportunities to explore, and desire to investigate and question science phenomena.	I enjoy exploring new activities about [*favorite topic inserted*] in school.	Litman and Spielberger (2003); Engelhard and Monsaas (1988); Kashdan et al. (2004)
Identity	The formation and role of identity in a child's experience with science is multifaceted. Our identity items focused on children's recognition of their role in science pursuits and their thoughts about themselves related to science and scientific pursuits.	I think like a science type person.	Girod (2009); Fraser (1981); Moore and Foy (1997)
Interest	As a cognitive-emotional construct, interest relates to people's affect toward science and the "predisposition to re-engage" in science. Interest is often argued to be key factor in science learning in terms of both engagement and deeper learning processes (e.g., finding connections in science	I would like to do activities related to robots at home.	Hidi and Renninger (2006); Germann (1988); Dawson and Bennett (1981); Dawson (2000); Renninger, Ewen, and Lasher (2002); Girod (2009)

(Continued)

TABLE 1
Continued

Construct	Conceptualization	Item Example	Citation Source Example
	content to one's own life, question generation) Items were constructed to ask about children's fascination with science, whether they actively seek out information on a science topic, and if they have a positive affect toward science and science topics.		
Persistence	Persistence can be conceptualized as actions taken to remain engaged when facing a difficult obstacle (e.g., a bad teacher, a failed experiment), or maintaining engagement in science activities over extended periods.	I would keep studying science, even if my teacher tells me I'm not good at it.	Duckworth and Seligman (2006); Lufi and Cohen (1987)
Responsibility	Children's responsibility is conceived as children's perception of their ability to organize science information, take an active part in their science learning, as well as examine their perceived control over their science learning.	When it comes to learning about [*favorite topic inserted*], having a good instructor is more important than how hard you try.	Niemiec, Ryan, and Deci (2010); Nowicki and Strickland (1973)
Expectancy value	Expectancy value is the hypothesized powerful combination of expectancy and value. A number of studies have found that when one has both the confidence in one's ability to successfully complete a task in addition to intrinsically or extrinsically valuing that task/content, one has very high motivation levels as displayed in a variety of output measures.	If I started a class project on climate change, I think I could do a really good job.	Eccles and Wigfield (2002); Nagengast et al. (2011)

rather to use them as a platform for understanding a child's overall preferences and motivation toward science along the dimensions in which science experiences vary. Items from existing theories were used to embed the tested dimensions (context, manner of interaction, topic). Throughout this study, we use "motivation" to refer to a child's inclination or desire to engage or participate in science experiences, as reflected in this variety of motivational constructs.

THE CURRENT STUDY

To explore and disentangle the potentially important variation in motivations due to these dimensions, we examined the relationship between children's motivation toward science across a range of experiences, varying systematically the manner of interaction with science, using different science topics, and referring to a range of places. We are interested in answering the main research question: *Does children's motivation shift along the dimensions of context, manner of interaction, and topic?* We hypothesized that children's responses about their motivation toward various science activities may be heavily influenced by these factors.

While many studies have demonstrated the importance of student science motivation on science achievement, less is known about how concrete science experiences relate to children's motivation or how these experiences build toward a child's developing understanding of science. Among the particular dimensions examined within this study, topic interest is likely the most robustly studied. Large-scale measurements have been conducted to examine science interest across particular topics (ROSE: Relevance of Science Education, PISA: Program for International Student Assessment); however, our current work offers notable important additions to this prior research. First, our students are at a developmentally younger age (11–12 years old) than the students in the ROSE and PISA data (15 years old) (ByBee & McCrae, 2011; Jenkins & Pell, 2006; Schreiner & Sjoberg, 2004), and our students are much older than the studies showing topic preferences at the start of formal education (e.g., Mantzicopoulos et al., 2008). These large age differences between our focus and the focus of prior work involve large changes in self-reflective thought, independence from adult supervision inside and outside school, social interactions with peers, as well as exposure and opportunities for science-related experiences. The intentional sampling of early middle school also affords us an opportunity to measure students' science motivation close to the start of the gradual decline in children's interest in science as they approach adolescence (e.g., Osborne et al., 2003; Simpson & Oliver, 1990; H. T. Zimmerman, 2012).

In addition, while topic interest is one major focus of this article, we also explore other dimensions of science motivation that are much less studied at any age, including exploring topic across context and manner of interacting. In a child's common experiences, there may be strong natural correlations among the dimensions such that some learning spaces or specific science domains lend themselves more easily to a specific manner of interaction. For example, perhaps science classrooms typically have less of a hands-on component and more reading and listening than do informal experiences. This example shows the potential overlap that may occur between dimensions, in this case context and manner of interaction. To assess the independent influences of each dimension, we balanced across these dimensions using a factorial design to understand the unique contributions of each dimension. In other words, questions about more active (e.g., "hands-on") experiences occurred with equal frequency in different contexts and within different science domains. Using this approach, we mitigated the potential problem of imbalanced dimensions by structuring our survey to measure dimension combinations in a more controlled, equal way.

TABLE 2
Participant Information Across Locations

		Age (Years)		
State	Testing Location	*M*	*SD*	Gender
California	71% museum	11.2	0.5	60% female
Pennsylvania	All school	11.4	0.6	58% female
Overall	31% museum; 69% school	11.3	0.5	59% female

METHOD

Participants and Recruitment

Two hundred and fifty-two fifth- and sixth-grade students from Pittsburgh, Pennsylvania, and the Bay Area, California, participated in the study (see Table 2 for description). All children in the Pittsburgh region were recruited through their school science classrooms, whereas Bay Area students were recruited through their school science classrooms or through their class visit to a local museum. Although student-level socioeconomic status (SES), ethnicity, and were not assessed, schools in both regions drew students from a range of SES and were not particularly higher or lower performing schools. From open online school enrollment data, Pittsburgh students are primarily Caucasian and African American and Bay Area students are largely Caucasian, Hispanic, and Asian. All students who were present on the day of survey administration completed the survey.

Materials

Survey: Topic Checklist. At the start of the online survey, children were asked which science topics they were interested in learning about from a list of science topics. This checklist was used to obtain a measure of children's interest at the topic level. The topic checklist included items sampled from five broad science disciplines: astronomy, biology, earth science, engineering, and physical science (e.g., astronomy was represented with "planets," "space travel," "telescopes," "distant galaxies," "The Moon," "The Sun," "black holes"). This combination of five disciplines with seven instances yielded 35 topics for the item checklist.

To ensure some basic level of familiarity and interest for late elementary-aged children, these 35 topic items were initially gathered from pilot testing conducted with fifth-grade students. As elementary school curricula are not currently standardized in the United States, children in the pilot testing were given a large list of science topics commonly learned in elementary school and found on other topic checklists in the literature. Children were asked which topics they found interesting and would like to learn more about as well as asked to generate their own list of science topics if they liked something that was not presented. As such, an in-depth knowledge of each topic was not necessary to make motivational judgments. More popular items were selected within each of the five broad science disciplines to produce seven items per discipline.

When children selected these topics at the beginning of the survey, they were instructed to select as many of the topics that interested them, but to pick a minimum of two. This checklist then generated two measures of topic interest (number of science topics each child selects and popularity of science domains and topics). Next, to measure maximal preferences as driven by a favorite topic, a list of the individual's selected topics was presented and each child was asked to select his or her one favorite topic.

TABLE 3
Example Items Labeled with Their Respective Dimension Coding

Example Items	Context	Manner of Interaction	Topic
I would like to *do activities* related to **robots** at home. (Interest item)	Informal	*Action*	**Engineering**
If I started a class *project* on **climate change**, I think I could do a really good job. (Expectancy value item)	Formal	*Action*	**Earth science**
I would keep *studying* **science**, even if my teacher tells me I'm not good at it. (Persistence item)	Formal	*Consuming new knowledge*	**General science**

Note. Underlined words of the items indicate context, italicized words indicate manner of interaction, and bold words indicate topic. Items simultaneously count toward one of each of the four dimensions.

Survey: Item Adaptation Along Dimensions. The remaining survey items consisted of 89 survey items asking about children's motivation and behavioral preferences toward science across the dimensions of context, manner of interaction, and topic. Please see the Appendix for a full list of items.

The selection of the seven motivational scales came from a national panel of researchers from cognitive, developmental, social, and educational psychology and science education convened to discuss key potential motivational constructs of relevance to late elementary that would be most predictive of long-term engagement in science; a goal was to look across theories rather than endorse any particular theoretical framework (Dorph, Schunn, Crowley, & Shields, 2011). Based on discussion of evidence and overlap of relevant literatures, these seven constructs were deemed likely be relevant to science motivation in late childhood/early adolescent development and not mutually redundant. Following input from the panel of experts, we conducted a series of pilot studies with the various subcomponents of the survey with fifth- and sixth-grade students to make sure the constructs were being measured reliably at this age. Edits were made to the survey based on this pilot, and then further adaptions were made to meet the research goals.

The items were constructed by adapting and extending existing motivational scales that have been previously argued to influence learning and engagement with science in formal and informal settings. These adaptations included adding a particular context, manner of interaction, or topic, when necessary (see Table 3). For example, "Everywhere I go, I am looking for new things or experiences." (Kashdan et al., 2004) was changed to "Everywhere I go, I am looking for new things about animals" to gain insight into students' topic interests. Some scales also needed to be adapted to be appropriate for late elementary rather than normed for college or high school (e.g., "I actively seek as much information as I can . . ." was changed to "I am often trying to find out more about . . .").

Context. Context was divided into formal science experiences (relating to school, classes, teachers), informal science experiences (at home, at a museum, with friends, at a camp), and a neutral category that did not specify a context (see Table 4 for categories within each dimension). Some items needed to be adapted to ask specifically about science (both formal and informal) rather than their original focus on other topics (e.g., "I think that what I am learning in this class is useful for me to know." (Pintrich & de Groot, 1990)

TABLE 4
Dimensions With Subscales and Number of Items

Context	Manner of Interaction	Topic	Motivation
Formal (27)	Consuming new knowledge (30)	Astronomy (stars) (7)	Appreciation (12)
Informal (27)	Analyzing (29)	Biology (plants) (7)	Curiosity (8)
Neutral (35)	Action (30)	Earth science (hurricanes) (7)	Identity (14)
		Engineering (robots) (7)	Interest (11)
		Physical science (gravity) (7)	Expectancy value (10)
		General science (38)	Persistence (18)
		Favorite (16)	Responsibility (16)
Total: 89	Total: 89	Total: 89	Total: 89

Note. Number in parentheses represents number of items in each dimension subscale. There are a total of 89 items in the survey, excluding the topic checklist. Each item fell into one category of the four dimensions simultaneously.

was changed to "What I know about science will be useful outside of school"); some scale items needed to be adapted to have a balance of locations for scales entirely focused on school or out of school (e.g., "I have a good feeling toward science" (Girod, 2009) was changed to "I have a good feeling when I think about science in school").

Manner of Interaction. Manner-of-interaction questions were also divided into three categories: consuming new knowledge, referring to the studying, reading, and going online for the learning of new science information; analyzing, which described a child's thinking about information they had previously learned; and action, specifying a hands-on activity.

Topic. The topic dimension included the same 35 items from the five broad science categories presented in the topic checklist, described above. These topics were embedded in items throughout the survey while maintaining an even distribution across our other dimensions of interest (e.g., context, manner of interaction). The remaining items were split into two categories: items asking about "science" at the general level ($n = 38$) to serve as a comparison for the topic items or items that were completed with the child's selection of their favorite subtopic from the topic checklist ($n = 16$). The "favorite" topic each child selected from the topic checklist was automatically inserted into specific survey items across the various subscales to ask about children's motivation and behaviors regarding their self-identified favorite science subtopic. For example, the item "When I am confused about——, I try and figure out an answer" was completed with each child's individualized response to their favorite item from the checklist.

Motivational Constructs. General personality/disposition scale items were modified to make them more specific tests for effects of context, manner of interaction, and topic (see Table 3). Appreciation and identity items were also modified for context and manner of interaction, but were kept at the science general level (e.g., "I am a person who thinks like a scientist") due to the manner in which appreciation and identity have been typically conceptualized. For example, asking about children's value at the science level (e.g., "Science is important to my daily life") made more conceptual sense than asking at the topic level with our topic instances (e.g., "Fossils are important to my daily life"). Similarly, identity is also

typically a science general subscale, examining the explicit knowledge of one's self related to science, and the items were more sensible when placed at the science general level (e.g., "I am a 'science' type person" vs. "I am a 'planet' type person"). Discipline-level terms are sensible (e.g., biologist or physicist), but we did not expect most children to be familiar with these terms.

Data Analysis Considerations Related to Our Measure

Previous research has shown that some of these scales tend to be correlated with each other, but measuring all these scales simultaneously is not common given that they originate from different motivational theories. We combine these measures to establish patterns across theoretical framings of motivation.

Rather than testing differences in means across motivational scales, the trends of children's positive or negative responses in response to context, topic, and manner of interaction were examined across motivational scales to understand children's preferences toward science. This decision was made because while it is mathematically possible to test potential mean differences between motivational scales in our survey (e.g., is interest higher than identity in this population?), the interpretation is difficult for a number of reasons. First, there is not necessarily a one-dimensional factor underlying each of the scales, as they contain meaningfully different dimensions that may be influencing answers. Second, two of the scales contained only general science items (appreciation and identity). Differences in means between motivational scales that varied across "general science" items and topic-based items may be driven by topic effects rather than true differences in levels across motivational constructs per se. Third, because the scale was not a full factorial across all dimensions, differences between motivational constructs could also vary on the popularity of the exemplars of each dimension (i.e., some science topics, such as "animals" were responded to more highly positive across topic list, favorite topic, and overall item mean).

Furthermore, because our survey has four orthogonal dimensions, traditional factor analysis at the raw item level cannot be used to extract (or confirm) any one dimension (e.g., motivation constructs). In addition, since we did not create items to represent the full factorial ($3 \times 3 \times 7 \times 7$) combinations of dimensions, it is also not possible to do factor analysis at intermediate aggregate levels because partial aggregates would be unbalanced (i.e., multiple dimensions were embedded within the same item, making it unclear which dimension was driving the factors). However, we do provide reliability indices for our measures of each dimension, which generally show high cohesiveness among items.

Procedure

All children completed the full 89-item survey in science class or during a class museum visit in a single sitting. The motivational scales (e.g., appreciation, persistence) were interspersed throughout the assessment to vary presentation of topic, manner of interaction, and context; but all children experienced the same order of questions. Children were asked to select the response that best represented how they felt about each item. Items were scored from –2 to 2 based on the following 5-point Likert scale: "YES!," "yes," "maybe," "no," NO!" and converted to *Z*-scores for analyses. The –2 to 2 scoring was used to make the numeric scale score meaningfully representative of the scale labeling. In other words, positive scale labels were coded with positive numbers (YES!, yes = 2, 1, respectively), and negative responses coded negative numbers (NO!, no = –2, –1, respectively; maybe = 0). All reversed items (e.g., "No matter how hard I try, I am confused by science") were reverse coded prior to analyses.

TABLE 5
Alphas, Means, and Intercorrelations of Motivational Subscales

Scale	a	M	SD	Curiosity	Identity	Interest	E-V	Persistence	Responsibility
Appreciation	.88	0.69	0.68	.70	.79	.75	.72	.76	.80
Curiosity	.74	0.56	0.71		.73	.77	.75	.78	.80
Identity	.91	0.39	0.78			.81	.77	.83	.84
Interest	.80	0.44	0.44				.80	.84	.84
Expectancy value	.77	0.56	0.56					.80	.81
Persistence	.88	0.47	0.47						.87
Responsibility	.83	0.36	0.36						

Note. N = 252. Mean scores range from –2 to 2. All correlations were significant at the $p < .001$ level. Column "E-V" represents the expectancy value.

RESULTS

Ruling Out Confounding Factors Between State and Testing Location

Comparative analyses on aggregate ratings were conducted to rule out potentially confounding effects of region (e.g., demographic differences associated with Pennsylvania vs. California) or testing context (within a museum vs. school) with the dimensions of focus here (i.e., formal/informal science preferences). No overall differences by region or testing context were found across any of the dimensions (e.g., children tested in the museum did not have higher informal question ratings than children tested in the school), and thus the testing location is not included within the analyses presented below.

Does Children's Motivation Shift Along the Dimensions of Context, Manner of Interaction, and Topic: Exploring Children's Sensitivity to Dimensions and Subscales

There are several important aspects to explore to effectively tease apart children's sensitivity and preferences to these various dimensions. Since the scale was a fractional factorial and not a full factorial through each of the dimensions (e.g., not every topic was placed in every context and in every manner of interaction), we examined these dimensions through correlation, main differences, and multidimensional scaling.

Motivational Subscales. Cronbach alphas for each of the seven scales ranged between $r = .74$ and .91, indicating that most of the scales could be adequately measured even with additional variance due to orthogonal manipulation of the other dimensions (as described below).

All the scales were highly correlated with one another ($r = .70$–.87, all significant at the $p < .001$; see Table 5). Paired-sample t-tests on these correlation coefficients revealed that responsibility and persistence subscales were most highly correlated with other motivational measures, curiosity was least correlated with other motivational measures, and the other measures correlating at intermediate levels.

Table 5 also displays the means for each subscale dimension, but as stated previously, formal testing between these means are neither inherently meaningful nor the goal of the current study. However, it is clear that children generally gave a modestly positive average response across all constructs, regardless of subscale, allowing for plenty of distance from scale end points to explore effects of context, topic, and manner of interaction.

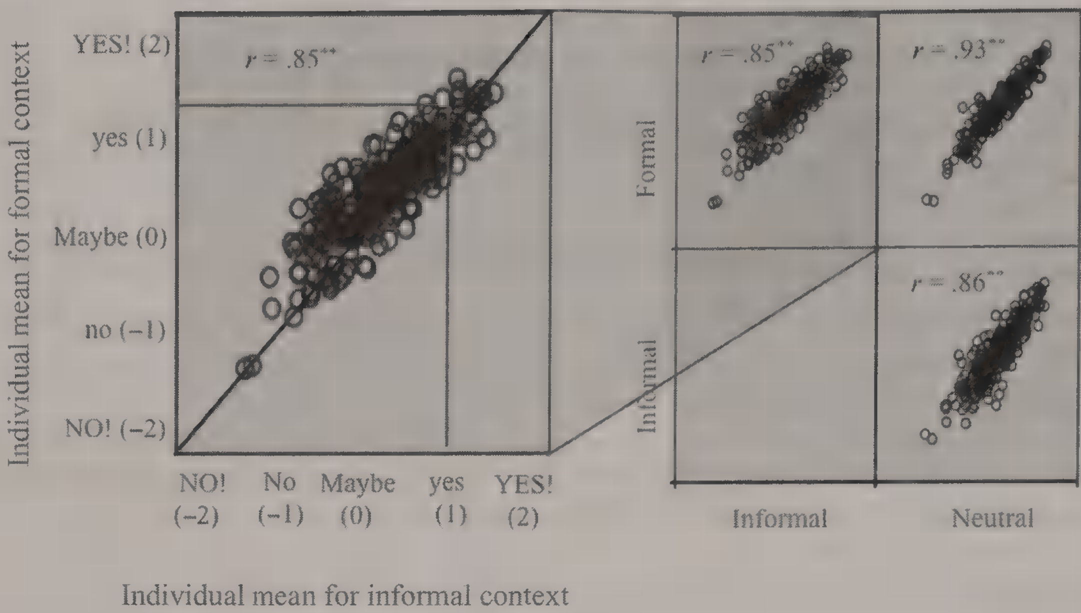


Figure 1. Correlations among context subscales.

Context. Items measuring context dimension (formal, informal, neutral) were used to generate a Crobach's alpha for each setting, respectively. In other words, the 27 items with formal context embedded in them were used to generate the reliability of formal items regardless of what motivational construct was being tested. With a large number of items per context, it was possible to produce high reliability in estimates of means for each context: The Cronbach alphas for each setting (formal, informal, and neutral) were very high (r = .91–.93). In contrast to the received wisdom that children should vary greatly in their responses to formal and informal science learning opportunities based on their highly individualized history of experiences in each context (e.g., a bad school experience or a lack of an informal learning experience), children's mean responses were remarkably consistent across context (r = .85–.93, all correlations significant at the $p < .001$ level). Figure 1 displays each of these correlations in more detail. Each dot represents the mean for a child plotted between two context subscales (e.g., mean responses for all informal context questions against the mean responses for all formal context questions). We can see that there are no large outlying cases (upper left corner or bottom right corner) in which a child responded consistently positively for one context and consistently negatively for another. Instead, we see that children tend to answer similarly on each (i.e., if they were strongly positive in their formal context responses; they were strongly positive in their informal context responses).

Since each context is roughly balanced across motivational constructs and the other dimensions of interest, comparisons across contexts are sensible. Examining the overall mean differences between subscales show that children tended to give more positive ratings to items related to the formal context or neutrally phrased items than to items related to the informal context ($t(251) = 11.72$, $p < .001$; $t(251) = 13.35$, $p < .001$, respectively; see Figure 2). This effect was moderate in size (Cohen's $d = 0.42$ between formal and informal and $d = 0.44$ between neutral and informal). Children's preference for formal context items is somewhat surprising given a common assumption discussed in the literature about the importance of informal science experiences for building a sense of fun versus formal science for building content knowledge (National Research Council, 2009). Conclusions drawn from such a finding should be interpreted carefully. For example, the differences may reflect aspiration rather than reality (e.g., I want to be interested in science experiences

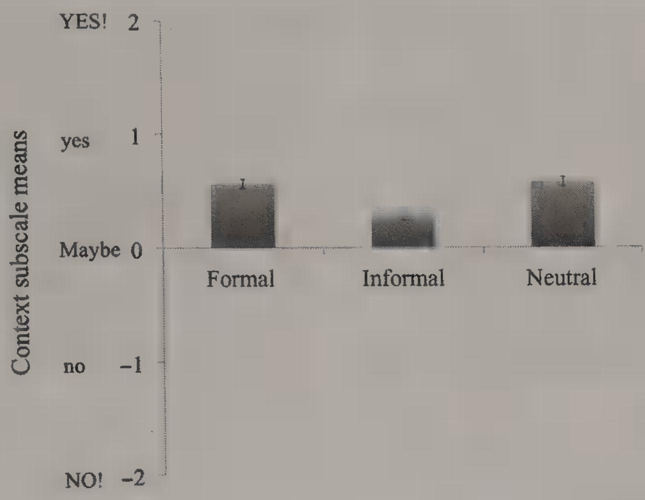


Figure 2. Average across context subscales.

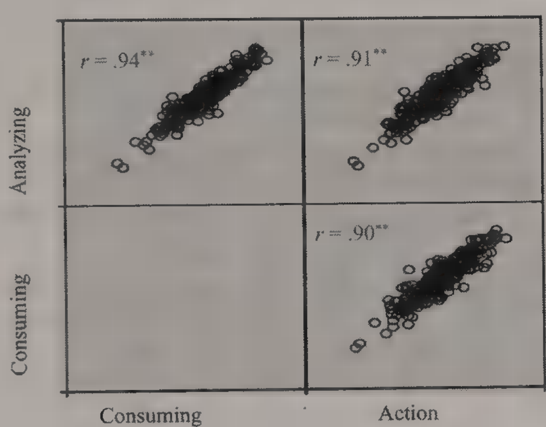


Figure 3. Correlations among manner-of-interaction subscales.

in school rather than I typically am interested in science experiences at school). Also, many children might have had relatively few prior informal science experiences, and the lack of experience might have driven down their agreement about motivational statements with respect to the informal context.

Manner of Interaction. As with the different contexts subscales, each manner-of-interaction subscale (consuming new knowledge, analyzing, action) had very high construct reliabilities ($r = .92–.93$). Correlations between each child's means for each manner of interaction were quite high ($r = .90–.94$, all correlations significant at the $p < .001$ level), showing that degree of preference for science is highly consistent across the manners of interaction (e.g., relative positivity toward analyzing items was very similar to relative positivity toward action). Figure 3 shows that no children had a very different relative response across the manners of interaction with science.

However, mean differences were observed across the manner-of-interaction items. A comparison of the subscale means showed that children responded less positively to items related to hands-on/action science activities, although this effect was small (analyzing and action: $t(251) = 6.77$, $p < .001$; consuming new knowledge and action $t(251) = 7.15$, $p < .001$; both differences had an effect size of $d = 0.19$; see Figure 4). There was no difference in responses between analyzing and consuming new knowledge ($t(251) = 0.53$, $p =$ n.s.). It is important to restate that each subscale was balanced across the other dimensions so that the difference found for action items is not due to confounds with

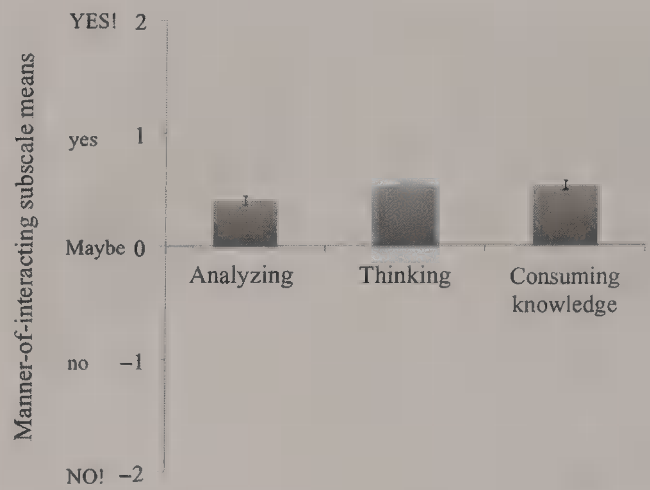


Figure 4. Correlations among manner-of-interacting categories.

TABLE 6
Intercorrelations of Topics

Topic Area	Item (*N*)	Biology	Earth Science	Engineering	Favorite	General Science	Physical Science
Astronomy	7	.53	.74	.60	.59	.72	.68
Biology	7		.65	.29	.51	.63	.53
Earth science	7			.55	.58	.76	.69
Engineering	7				.45	.58	.63
Favorite	16					.67	.61
General science	38						.79
Physical science	7						

Note. $N = 252$. All correlations were significant at the $p < .001$ level.

informal context items. The small preference against action items occurred for questions about the informal and for questions about the formal contexts.

Topic. In contrast to the small variation in preferences associated with the previous dimensions, children showed considerable differentiation by topic. There were fewer items per topic than per context or manner of interaction and the alphas for topics were somewhat lower, ranging from $r = .61$ to $.80$. In addition, there were significant, but smaller and more varied correlations between broad topic areas, ranging from $r = .29$ to $.79$ (see Table 6). Figure 5a presents the largest divergence by topic (biology against engineering). In this figure, we see many children's means plotted in the upper left and lower right quadrant, instances of a child responding positively toward one topic (e.g., biology), but negatively toward another (e.g., engineering). Thus, children did respond to topics with considerably more differentiation than they did to contexts and manner of interaction. Figure 5b also shows that children differentiated between topics (e.g., biology) and items asking about science at the general level.

Comparisons of the means across topic subscales shows us that children varied in their overall preferences across topics, with their favorite topic receiving the highest positive response. Biology, physical science, and engineering subscales received similar responses from children in that there were no significant mean differences between them. Astronomy and earth science were not different from each other, but were each, respectively, different

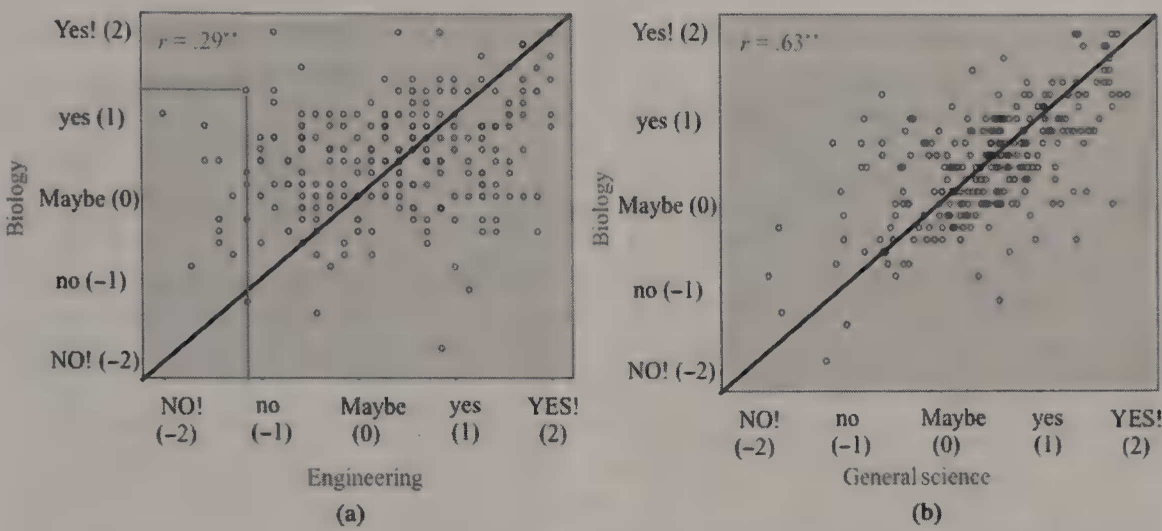


Figure 5. (a) Example of divergence between biology and engineering and (b) correlation between biology and general science topics.

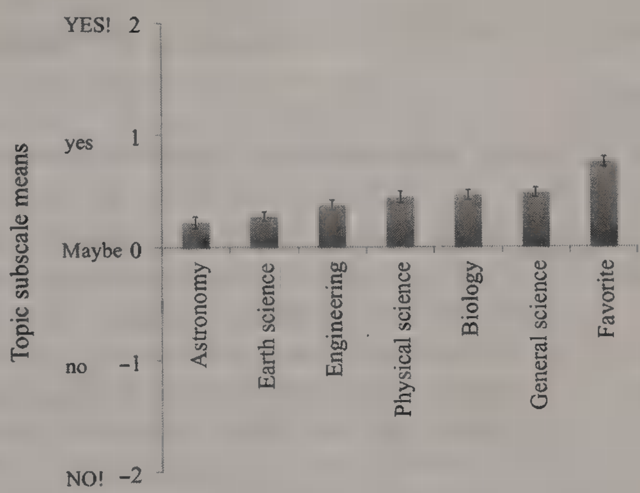


Figure 6. Averages across topic subscales.

from biology, physical science, and engineering (see Figure 6). That is, while engineering and biology are the most differentiated by individuals, the overall means are similar, suggesting that some kids strongly prefer engineering over biology whereas other kids strongly prefer biology over engineering.

Multidimensional scaling (MDS) was conducted and a three-dimensional scale provided a good to fair fit (Stress I = 0.09) and echoed the pattern above, with biology, physical science, and engineering clustering more closely together than astronomy and earth science (see Figure 7a).

Children's mean responses to the 32 items containing general science were similar to biology and physical science means, but varied from the engineering, astronomy, and earth science means. When placed into the MDS analysis, these general science items were more related to biology, physical science, and earth science more than astronomy and engineering, although the inclusion of general science did raise the Stress I slightly (Stress I = 0.12). Most critically, while there were generally moderately strong correlations between the general science means and the means on the other questions, assessments via questions about "science" are not synonymous with questions using more specific topics, like various biology topics (see Figure 7b).

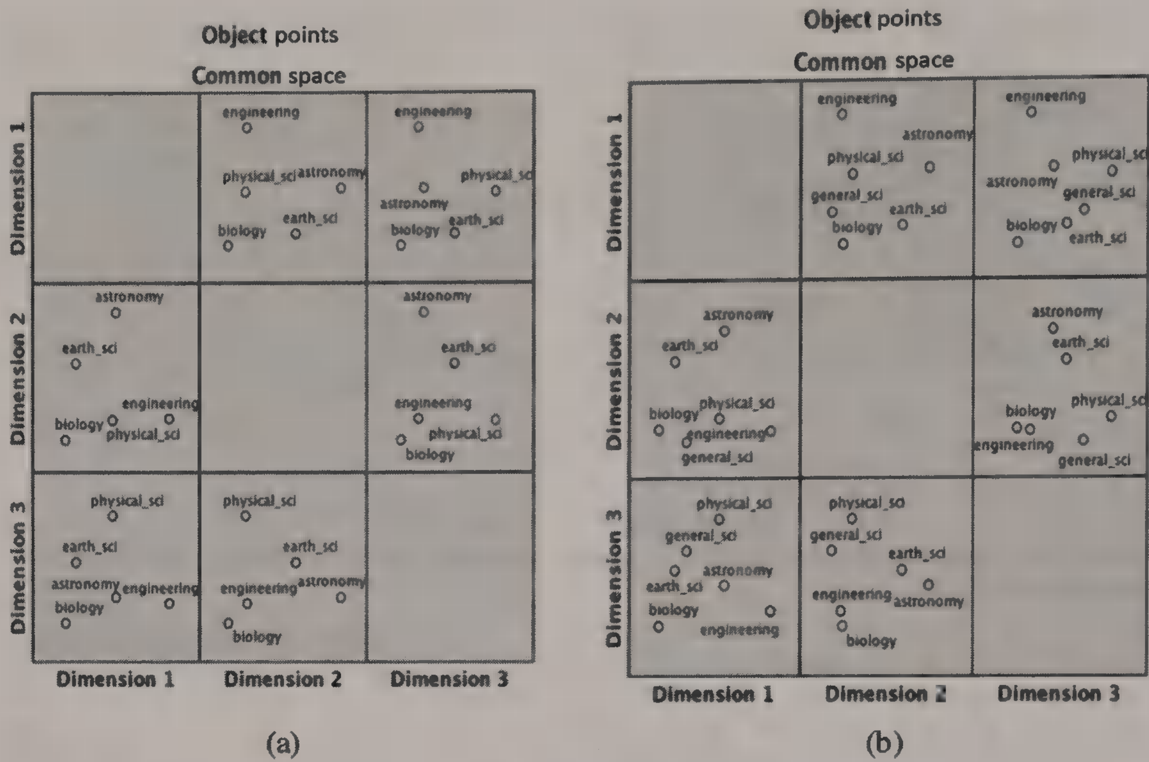


Figure 7. (a) MDS results with general science omitted and (b) MDS results with general science included.

Topic-Specific Preferences. Overall topic preferences can also be explored via the topic checklist data. Children chose an average of 10.1 items from the checklist ($SD = 7.8$), with a range of 1–35. Over 60% of children selected 10 items or less and roughly 80% selected 15 items or less, giving the distribution a rather positive skew. When examining which science domains (e.g., biology) were most popular, a similar pattern to the survey items was found, with biology items being most popular ($M = 0.32$, $SD = 0.26$), followed by physical science ($M = 0.32$, $SD = 0.27$), and engineering ($M = 0.29$, $SD = 0.30$). Earth science and astronomy again were slightly lower than other subscales ($M = 0.26$, $SD = 0.28$; $M = 0.26$, $SD = 0.29$), respectively. Within science domains, the most popular items did not represent a single domain, but were spread out across the different categories. Biology, earth science, physical science, and engineering were all represented in the five most popular items (animals, sea life, crystals, robots, oceans).

Children's favorite item selection also followed a similar subscale ranking. Thirty-eight percent of children selected a biology topic as their favorite followed by, physical science and engineering (each with 22%), astronomy (10%), and earth science (8%). The top 10 selected items were all chosen at least 10 times and resulted in the following most commonly favorite items: animals, sea life, optical illusions, robots, computers, DNA, crystals, body systems, technology, and chemicals. We see less astronomy and earth science preference, as was found throughout the Likert ratings.

DISCUSSION

The data presented here help increase our understanding of environmental features that shape children's motivation for science learning. We examined this motivation at a crucial time of development (the beginning of adolescence) when children's choice and autonomy generally increase (Wray-Lake, Crouter, & McHale, 2010). Gaining insight into the dimensions that ignite, support, and maintain children's science motivation during this

time aids us in discovering ways to encourage such motivation. Most saliently, we find that specific science content has the largest effects. In addition, when means by context and manner of interaction did vary, they did so in surprising ways.

Context and Manner-of-Interaction Effects

Children's lower preference for informal activities and more active science opportunities is somewhat surprising considering many of these activities offer a higher degree of autonomy and freedom from graded assessments, and such freedom has been shown to increase intrinsic motivation (Black & Deci, 2000). However, the fact that an activity is "informal" or "hands on" may not be enough to motivate children effectively, as a number of prior studies have found (Areepattamannil, 2012; Klahr, Triona, & Williams, 2007). Since topic interest is intricately related to children's motivation, developing content-related activities that cross different manners of interaction may be beneficial. However, it seems that solely altering the context or the way a child interacts with material may not be enough to ignite their engagement. It could be that topic is such a large driver of interest and motivation that these other dimensions matter very little; for example, it could be that even a very didactic presentation on a highly interesting topic is more engaging to a child than a very rich task about a topic that is very uninteresting to that child.

Our scope of formal and informal contexts was broad to capture a wide range of common childhood science activities. There may be more fine-grained differences within these categories that may produce greater differentiation by context or manner of interaction. For example, formally testing whether social involvement with peers, learning at home versus at a museum, or children's participation in past science experiences (e.g., many vs. few science experiences; many hands-on experiences vs. few hands-on experiences) moderates preferences could be done with a survey instrument focused on these dimensional differences and the potential interactions among them. Yet, we should be careful to consider that these dimensions are somewhat blurry in real-world situations and children will engage in a wide range of science activities that vary dynamically and not always consistently (Dierking et al., 2003). We would lack full understanding of a child's science experience if we were to mistake these clear distinctions as the concrete ways in which children explicitly break up their perceptions of science. However, as we have seen in our data, children do hold different overall preferences toward contexts that can be useful to consider in design and implementation of science activities and challenge the assumption that informal activities are always more motivating.

Discipline and Topic Effects

Discipline Preferences. Children showed greatest sensitivity and preference to variations in science disciplines. Biology and physical science topics were most popular, and earth science and astronomy were least popular, as has been found in previous research with younger ages (Mantzicopoulos & Patrick, 2010). Our analyses show that some of these domains appear to be more closely associated than others, such as earth science with astronomy; however, how the children's motivation toward particular domains can be explored further in our data.

Topic Preferences. Our topic-based approach allowed us to examine popular topics within and across these larger domains. First, we find that children have specific interests at the individual topic level (e.g., robots) far beyond domain-level preferences (e.g.,

engineering). In other words, some topics were overall much more popular than others (e.g., animals, sea life, crystals, robots), irrespective of their larger domain category. For example, earth science was one of the least favored domains, yet "oceans" ranked as one of the most interesting topics for children. Second, even when asked at the individual topic level (i.e., the checklist), children reported a range of interest in various science topics, as is evidenced by their average selection of about 10 topics from the general checklist. By study design, the selection of 10 topics exceeds what may be found in one domain, showing many children held interests beyond one popular domain. How these topics and domains interrelate gives us insight into how children categorize and group their science interests at this age.

Alternative Topic Clustering. Our current grouping of science topics was based on common boundaries of what composes these larger science domains. As such, we placed topics such as "plants" in biology and "telescopes" in astronomy. However, at this early age, children's conceptualization of these domains is unlikely to be so well distinguished. Children are still familiarizing themselves with different sciences and science content as they progress in their learning and experience. Understanding the boundaries between domains may not be obvious or even clearly stated. Science experiences and curricula may not have clearly addressed "physics" as a distinct field, for example. We examined children's responses in a top-down approach, looking to see whether canonical categorizations of science categories emerged. If children were aware of their overall biology interest, they may be more likely to pick topics relating to this overall field. However, children likely considered each topic rather independently, given their age and experience with science material, especially given the random presentation of topics over the 89 items. While our data show Cronbach alphas for each science domain were moderately strong, children's interests may still pull from a variety of domains.

In fact, children's preferences may transcend the typical boundaries of science domains and even form different clusters. Perhaps, in their science experiences, some concepts are more associated than others. For example, a child may first learn about "gravity" in the context of a lesson on the planets. Although "gravity" is grouped as a physical science concept, perhaps it more closely aligns with "planets," "stars," and "the Moon" (*astronomy*) in the minds of children. Previous work has raised questions about the perception of science categories, positing that teenagers may express interest in topics for reasons other than their domain content (Jenkins & Pell, 2006). ByBee and McCrae (2011) found that adolescent males tended to express higher levels of interest in topics that have a technological component, even if the topic is not directly related to technology (e.g., pollution). This example shows one way in which science domains are not straightforward in children's minds, but can vary due to another dimension, such as procedural methods. Further work across different children's development would help us understand the composition of these dimensions at various ages, and how this structure may change through a child's experience with different kinds of science.

ByBee and McCrae's (2011) finding also raises an additional consideration: There are inherent procedural differences among science domains. While we attempted to constrain these differences as much as possible in our assessment, there are necessary variations in these disciplines beyond content that involve the way research is conducted, the distance from the object being studied, the speed of return of research, how this research is communicated, and how socially interactive the research field may be. For example, robots, chemical experiments, and electrical circuits all have a very physical and mechanical element. The result of such endeavors often yields a rather physical, occasionally immediate

payoff (e.g., circuit works and a light comes on). Examining the habitat of an endangered animal requires a different set of tools and engagement resulting in outcomes that may seem differentially rewarding to various children (e.g., acidity values in water are lowered). Certainly there is overlap in the scientific thinking and inquiry between both, yet they seem qualitatively different in some of the processes. Perhaps some of the differences in science preference could be explained by the opportunities different sciences allow their scientists. Children at this age may not be fully aware of the details of such differences, but may be drawn to different science activities due to such variables. More exploratory studies focusing on children's perceptions of the relationships between sciences and their subtopics may provide further insight into children's science preferences.

Implications

Measurement. One goal of this research is to raise awareness of the value of measuring children's science perceptions at the topic level, in addition to research asking about science at a general level. While research has become increasingly domain specific (e.g., Eccles & Wigfield, 2002), domain specificity still allows for large degrees of unaccounted variation. A topic-based approach allows researchers to explore breadth of science interest (how many topics interest children?), the salient topics that are popular at various developmental ages (which topics interest children?), how these dimensions affect achievement and outcome variables, and how design implementation can improve children's interest and engagement. In addition, topic-level items allow researchers to probe children's motivations toward science more implicitly. Children vary in their preferences for science activities that they may not personally identify as science. Insightful qualitative work has shown this to be the case (Bell, Brickner, Lee, Reeve, & Zimmerman, 2006); larger scale assessments may miss cases of deep science interest if we skim the surface of children's preferences based on their interpretation of the word "science" itself.

Whether a scale examines motivation at the topic or at the general science level largely relies on the research question being asked. Our data show that responses at the general science level correlate with the aggregate using all the topic items at $r = .84$, demonstrating relatively good relationship between a child's overall awareness of their motivations toward "science" and their topic-based interests. However, there remains a trade-off between these two approaches that should be thoughtfully considered when selecting a method. For all the benefits of a topic-based approach, topic-specific surveys require more items to generate an approximation of children's motivation toward a domain within a specific theory. This raises questions about methodological constraints, such as length of test and the ordering presentation of items. It may also not answer questions about children's overall perception of science and their relationship to it; as much of the early learning environment comes with the label "science" (e.g., the Carnegie Science Center, sixth-grade science class, Sid the science kid), relationship with that label is important.

Alternatively, science general surveys are useful for answering a variety of questions and are an appropriate method for examining children's motivation in science, but their shortcomings should also be recognized and their application and generalizations should be carefully considered before administration. Asking science general questions forces ambiguity on the respondent when they like some aspects of science but not others. How individuals choose to handle that (some responding as if the question is asking about their favorite topic only vs. some asking about all topics) will inevitably be varied and thus lead to measurement error. Other differences occurring in subgroups, such as gender, vary greatly across science topics (Jones, Howe, & Rua, 2000; Tyson et al., 2007) and may be obscured at the science general level. Depending on a researcher's line of questioning, these

may or not be highly influential to the question at hand, and researchers should decide what is most appropriate for their purposes, acknowledging these trade-offs.

Interventions. In alignment with the idea that science learning is cumulative throughout a child's life (Dierking et al., 2003), it is beneficial to help children connect their learning experiences across various contexts. These connections are not always spontaneously made or obvious to children (Stake & Mares, 2005). Helping them become more aware of ways to find, engage, and connect their curiosity and interest across different settings may help deepen their knowledge and persistence in science learning (Hofstein & Rosenfeld, 1996; Stake & Mares, 2005). There are many ways this could be enacted, and intervention and design-based research could help direct concrete future steps. Our purpose here was to examine children's sensitivity in motivation toward different science dimensions that could help inform this work.

With forthcoming work demonstrating a connection among items used in our survey and student engagement and choices in science learning (Sha et al., 2013), research exploring children's preferences toward the different dimensions of science may inform future development of early science activities (e.g., topical summer camp program), showing us where to focus to most effectively meet student interest and value toward science content and processes. It should be noted, however, that our work here focuses on how various situational aspects shape student motivation, but not which features shape learning outcomes. Consideration of student motivation, as well as learning outcomes, should be considered when developing science activities.

Research has shown that the importance of generating situation interest, regardless of topic, can help student engagement and learning (Hidi & Harackiewicz, 2000; Jarrett, 1999). As such, educators should feel encouraged to help scaffold students' potential interest in a topic they have yet to find motivating. However, in free-choice learning situations (camps, after school programs, elective courses), situational interest cannot be triggered if children choose not to come at all, and understanding children's topical interest can influence choices that will maximally recruit additional learners. For example, teaching and out-of-school science experiences could focus on specific topics that broadly appeal to children overall, or specifically at different developmental ages (Trumper, 2006a, 2006b). While children's differences in topic interest may appear to make topic selection more difficult for educators, clear trends emerge that can help direct content choices and development. Our data suggest that some combination of biology and engineering content may easily capture the interest of most children, specifically topics around animals, robots, and computers. Other, less inherently interesting topics would need to be introduced in a way that engages students to support the development of interest in those topics, for example, through consideration of important applications.

Considerations for Motivational Variables. Rarely are many motivational theories measured simultaneously and therefore not a great deal is known about their relationships among constructs across theories. The high correlation between these variables could mean a number of things. Perhaps some of these constructs co-occur within an individual (e.g., expectancy value and identity) and are part of an underlying latent factor that explains the relationship. Alternatively, some of the correlation among the variables may be due to lower metacognitive awareness in children at this age (Veenman, Van Hout-Wolters, & Afflerbach, 2006; Whitebread et al., 2010). Specific comparisons among motivational theories were not the focus of our current study, yet the considerably high correlation among them is worth noting. Do we know how these theories interact or relate to each other?

The high intercorrelations suggest that correlational findings in favor of one theory may also have produced correlational findings in favor of other theories as well. However, as we may have partially disrupted the typical relationships due to the embedding of other dimensions in these items, we cannot definitively say we are measuring each motivational construct distinctly, but rather are sampling a broad range of motivations. Future research should consider this positive manifold among these different theory-inspired motivational measures in more depth to clarify their coexistence within an individual (Pintrich, 2003).

APPENDIX: SURVEY ITEMS ORGANIZED BY CONSTRUCT

Appreciation
Thinking about science is important to my life.
All people should learn lots of science in school.
It's important to be good at doing science in order to get a good job.
Understanding science helps people make sense of today's world.
Scientists cause more good than bad in the world.
Scientists make our lives better.
Scientific theories change all the time.
Understanding science is helpful for solving problems.
Science can solve nearly all problems.
Most people should visit a museum to think about science.
What I know about science will be useful outside of school.
My science class will make me a better thinker.
Curiosity
Outside of science class, I often wonder about global warming.
I am curious to learn how the body works.
I like to mess around with new technology.
I enjoy exploring new activities about____in school.[a]
It is cool to learn new things about gravity in school.
Everywhere I go, I am looking for new activities about____.
Wherever I go, I am interested in discovering new facts about____.
I get excited about discussing space in school.
Interest
I would like to learn more about hurricanes in school.
I often watch TV shows and/or read about space travel.
I would like to look closely at fossils in a museum.
In school, thinking about topics like molecules makes me yawn.
Sometimes thinking about____is boring to me.
I have a good feeling when I do science activities in school.
I often think about science topics at home.
Thinking about DNA is interesting to me.
I feel good when I learn about optical illusions in school.
I use the internet to find information about____.
I would like to do activities related to robots at home.
Expectancy Value
I like to learn new facts about black holes by watching TV shows.
Learning about sea life is important to me.

When I'm confused about____, I try to figure out an answer.
If I started a class project on climate change, I think I could do a really good job.
I want to learn everything about____, even if it's complicated.
If I attend a science camp, I would expect that my project would be the best.
I would go to a summer camp to build a solar energy project.
It's important to me to be an expert using computers in school.
I am afraid I will do a bad job learning about____in school.
I know I can learn a lot about electricity.
Persistence
I would think about magnets in school over and over until I understood them.
I would use my free time at school to put extra effort into a volcano activity.
I am OK with thinking about____even if I don't understand it at first.
I'm ok with trying again if a model rocket activity doesn't work at first.
If I watched a TV show about the moon, I would keep thinking about it even after the show was over.
I would build a science project at camp, even if none of my friends are interested in science.
I would like to spend lots of time looking at stars through a telescope in my back yard.
In school, I would keep thinking about how crystals form, even if it was hard.
I would continue watching a TV show about science even if it gets confusing.
When I am thinking about a science problem, I keep going until I understand.
I will keep doing a class activity about the ocean, even if I have to keep at it for a long time.
I need people to cheer me on to keep working on activities about plants.
If I have started an activity about bugs and butterflies at home and it seems like it is going to take a long time, I will stop doing it.
I would keep reading a book about science even if it was hard or long.
I would never choose to do an activity about the sun that takes more than a few hours.
I would keep studying science, even if my teacher tells me I'm not good at it.
I would study science even if I have a bad teacher.
I would spend my free time learning about____even if my parents do not think it is important.
Responsibility
I can learn about ecosystems in school if I try hard enough.
If I'm having trouble thinking about science in school, working harder can make a big difference.
When it comes to learning about____, having a good instructor is more important than how hard you try.
I would take out a library book about science.
I would ask my parents to take me to the zoo to learn about animals.
I know who to ask if I want to know more about planets.
I often make time to think about____outside of school.
I'm able to get information on mixing chemicals from the web on my own.
I would ask my parents to let me attend a camp where we build and test structures.
I get science projects done without my teacher or parents telling me to.
To think like a scientist, you have to have a special talent.
With enough time, I could learn science in school.
I enjoy discussing what I know about____with other people.
I want to help people think scientifically.
I would try taking apart an old computer at home by myself.
I always look forward to talking to my friends about earthquakes.

Identity
I think like a science type person.
Other people think I'm good at doing science.
I am the type of person who could work as a scientist someday.
Learning about____would be very easy for me in school.
No matter how hard I try, I am confused by science. (R)
I often think, "I will fail" when a science activity seems hard. (R)
I am bad at doing science activities. (R)
When I think about the word "science," I have a bad feeling. (R)
I feel uncomfortable when other kids talk to me about science. (R)
I have a good feeling when I think about science in school.
It is important for me to learn about____over summer vacation.
I am a person who thinks like a scientist.
I often investigate____so that I can understand how things work.
I often investigate science in my free time so that I can learn more about it.

[a]A blank space ("____") indicates that a child's self-selected favorite topic item was inserted automatically into the item via the survey system.

REFERENCES

Adams, R., Evangelou, D., English, L., Dias de Figueiredo, A., Mousoulides, N., Pawley, A. L., et al. (2011). Multiple perspectives on engaging future engineers. Journal of Engineering Education, 100(1), 48–88.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., WIllis, B., & Wong, B. (2012). Science aspirations, capital, and family habitus: How families shape children's engagement and identification with science. American Educational Research Journal, 49(5), 881–908.

Areepattamannil, S. (2012). Effects of inquiry-based science instruction on science achievement and interest in science: Evidence from Qatar. Journal of Educational Research, 105(2), 134–146.

Bell, P., Brickner, L. A., Lee, T. R., Reeve, S., & Zimmerman, H. T. (2006). Understanding the cultural foundations of children's biological knowledge: Insights from everyday cognition research. Presented at Proceedings of the Seventh International Conference of the Learning Sciences, Bloomington, IN.

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. Science Education, 84, 740–756.

Bryan, R. R., Glynn, S. M., & Kittleson, J. M. (2011). Motivation, achievement, and advanced placement intent of high school students learning science. Science Education, 95(6), 1049–1065.

ByBee, R., & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. International Journal of Science Education, 33(1), 7–26.

Crowley, K., & Jacobs, M. (2002). Building islands of expertise in everyday family activity. In K. Crowley & K. Knutson (Eds.), Learning conversations in museums (pp. 333–356). Mahwah, NJ: Erlbaum.

Dawson, C. J. (2000). Upper primary boys' and girls' interests in science: Have they changed since 1980? International Journal of Science Education, 22(6), 557–570.

Dawson, C. J., & Bennett, N. (1981). What do they say they want? Year 7 students' preferences in science. Research in Science Education, 11, 193–201.

Dierking, L. D., & Falk, J. H. (2003). Optimizing out-of-school time: The role of free-choice learning. New Directions for Youth Development, 97, 75–88.

Dierking, L. D., Falk, J. H., Rennie, L., Anderson, D., & Ellenbogen, K. (2003). Policy statement of the "informal science education" ad hoc committee. Journal of Research in Science Teaching, 40(2), 108–111.

Dorph, R., Schunn, C. D., Crowley, K., & Shields, P. (2011). An agenda for investigating and investing in effective science learning opportunities for young learners: The science learning activation lab. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, LA.

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. Journal of Educational Psychology, 98(1), 198–208.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. Annual Review of Psychology, 53, 109–132.

Engelhard, G., & Monsaas, J. A. (1988). Grade level, gender, and school-related curiosity in urban elementary schools. Journal of Educational Research, 82(1), 22–26.

Eshach, H., & Fried, M. N. (2005). Should science be taught in early childhood? Journal of Science Education and Technology, 14(3), 315–336.

Fraser, B. (1981). TORSA: Test of science related attitudes. Hawthorne, Victoria: Australian Council for Education Research.

Fulp, S. L. (2002). 2000 National survey of science and mathematics education: Status of electuary school science teaching. Retrieved June 27, 2012, from http://2000survey.horizon-research.com/reports/elem_science/elem_science.pdf.

Germann, P. J. (1988). Development of the attitude toward science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. Journal of Research in Science Teaching, 25(8), 689–703.

Girod, M. (2009). Comparing the added value of blended science and literacy curricula to inquiry-based science curricula in two 2nd-grade classrooms. Journal of Elementary Science Education, 21(3), 13–32.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for 21st century. Review of Educational Research, 70(2), 151–179.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. Educational Psychologist, 41(2), 111–127.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. Science Education, 88(1), 28–54.

Hofstein, A., & Rosenfeld, S. (1996). Bridging the gap between formal and informal science learning. Studies in Science Education, 28, 87–112.

Jacobs, J. E., Finken, L. L., Griffen, N. L., & Wrightm, J. D. (1998). The career plans of science-talented rural adolescent girls. American Educational Research Journal, 35(4), 681–704.

Jarrett, O. S. (1999). Science interest and confidence among preservice elementary teachers. Journal of Elementary Science Education, 11(1), 47–57.

Jenkins, E. W., & Pell, R. G. (2006). The relevance of science education project (ROSE) in England: A summary of findings. Leeds, England: Centre for Studies in Science and Mathematics Education, University of Leeds.

Jones, M. G., Howe, A., & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. Science Education, 84(2), 180–192.

Kashdan, T. B., Rose, P., & Fincham, F. D. (2004). Curiosity and exploration: Facilitating positive subjective experiences and personal growth opportunities. Journal of Personality Assessment, 82(3), 291–305.

Klahr, D., Triona, L. M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. Journal of Research in Science Teaching, 44(1), 183–203.

Lent, R. W., Brown, S. D., & Larkin, K. C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. Journal of Counseling Psychology, 31(3), 356–362.

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. Journal of Personality Assessment, 80(1), 75–86.

Lufi, D., & Cohen, A. (1987). A scale for measuring persistence in children. Journal of Personality Assessment, 51(2), 178–185.

Mantzicopoulos, P., & Patrick, H. (2010). "The seesaw is a machine that goes up and down": Young children's narrative responses to science-related informational text. Early Education & Development, 21(3), 412–444.

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2008). Young children's motivational beliefs about learning science. Early Childhood Research Quarterly, 23(3), 378–394.

Mantzicopoulos, P., Samarapungavan, A., & Patrick, H. (2009). "We learn how to predict and be a scientist": Early science experiences and kindergarten children's social meanings about science. Cognition and Instruction, 27(4), 312–369.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. Behavioral and Brain Sciences, 34(2), 57–111.

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. Review of Educational Research, 65(2), 93–127.

Moore, R. W., & Foy, R. L. H. (1997). The scientific attitude inventory: A revision (SAI II). Journal of Research in Science Teaching, 34(4), 327–336.

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K. T., & Trautwein, U. (2011). Who took the "x" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. Psychological Science, 22(8), 1058–1066.

National Research Council. (2009). Learning science in informal environments: People, places, and pursuits. P. Bell, B. Lewenstein, A. W. Shouse, & M. A. Feder (Eds.). Washington, DC: The National Academies Press.

Niemiec, C. P., Ryan, R. M., & Deci, E. L. (2010). Self-determination theory and the relation of autonomy to self-regulated processes and personality development. In R. H. Hoyle (Ed.), Handbook of personality and self-regulation. Oxford, England: Wiley-Blackwell.

Nowicki, S., & Strickland, B. R. (1973). A locus of control scale for children. Journal of Consulting and Clinical Psychology, 40(1), 148–154.

Ntoumanis, N., Taylor, I. M., & Thogersen-Ntoumani, C. (2012). A longitudinal examination of coach and peer motivational climates in youth sport: implications for moral attitudes, well-being, and behavioral investment. Developmental Psychology, 48(1), 213–223.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. International Journal of Science Education, 25(9), 1049–1079.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. Journal of Educational Psychology, 95(4), 667–686.

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology, 82(1), 33–40.

Renninger, K. A., Ewen, L., & Lasher, A. K. (2002). Individual interest as context in expository text and mathematical word problems. Learning and Instruction, 12, 467–491.

Schmidt, W. H., McKnight, C., & Raizen, S. (1997). A splintered vision: An investigation of U.S. science and mathematics education. Dordrecht, The Netherlands: Kluwer.

Schreiner, C., & Sjoberg, S. (2004). Sowing the seeds of ROSE. Background, rationale, questionnaire development and data collection for ROSE (The Relevance of Science Education)—A comparative study of students' views of science and science education. (4/2004), Oslo, Norway: Department of Teacher Education and School Development, University of Oslo.

Sha, L., Schunn, C. D., & Bathgate, M. E. (2013). The roles of self-efficacy, and effort regulation dispositions of self-regulated learning in predicting choice and engagement in science learning. Manuscript submitted for publication.

Simpson, R. D., & Oliver, J. S. (1990). A summary of major influences on attitude toward and achievement in science among adolescent students. Science Education, 74(1), 1–18.

Stake, J. E., & Mares, K. R. (2005). Evaluating the impact of science-enrichment programs on adolescents' science motivation and confidence: The splashdown effect. Journal of Research in Science Teaching, 42(4), 359–375.

Trumper, R. (2006a). Factors affecting junior high school students' interest in biology. Science Education International, 17(1), 31–48.

Trumper, R. (2006b). Factors affecting junior high school students' interest in physics. Journal of Science Education and Technology, 15(1), 47–58.

Tyson, W. (2011). Modeling engineering degree attainment using high school and college physics and calculus coursetaking and achievement. Journal of Engineering Education, 100(4), 760–777.

Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. Journal of Education for Students Placed at Risk, 12(3), 243–270.

Vadeboncoeur, J. A. (2006). Engaging young people: Learning in informal contexts. Review of Research in Education, 30(1), 239–278.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. Metacognition and Learning, 1, 3–14.

Vygotsky, L. S. (1978). Mind and society: The development of higher mental processes. Cambridge, MA: Harvard University Press.

Weinburgh, M. H., & Steele, D. (2000). The modified attitudes toward science inventory: Developing an instrument to be used with fifth grade urban students. Journal of Women and Minorities in Science and Engineering, 6, 87–94.

Whitebread, D., Almeqdad, Q., Bryce, D., Demetriou, D., Grau, V., & Sangster, C. (2010). Metacognition in young children: Current methodological and theoretical developments. In A. Efklides & P. Misailidi (Eds.), Trends and prospects in metacognition research (pp. 233–258). Dordrecht, The Netherlands: Springer.

Wigfield, A., Guthrie, J. T., Tonks, S., & Perencevich, K. C. (2004). Children's motivation for reading: Domain specificity and instructional influences. Journal of Educational Research, 97(6), 299–309.

Wray-Lake, L., Crouter, A. C., & McHale, S. M. (2010). Developmental patterns in decision-making autonomy across middle childhood and adolescence: European American parents' perspectives. Child Development, 81(2), 636–651.

Zimmerman, C. (2000). The development of scientific reasoning skills. Developmental Review, 20(1), 99–149.

Zimmerman, H. T. (2012). Participating in science at home: Recognition work and learning in biology. Journal of Research in Science Teaching, 49(5), 597–630.

Science
Education

Expectancy-Value Models for the STEM Persistence Plans of Ninth-Grade, High-Ability Students: A Comparison Between Black, Hispanic, and White Students

LORI ANDERSEN, THOMAS J. WARD
Department of Educational Policy, Planning, and Leadership, The College of William and Mary, Williamsburg, VA 23185, USA

Received 27 November 2012; accepted 9 October 2013
DOI 10.1002/sce.21092
Published online 27 November 2013 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Group differences in the effects of the expectancies and values that high-ability students have for science and mathematics on plans to persist in science, technology, engineering, and mathematics (STEM) were investigated. A nationally representative sample of ninth-grade students, the High School Longitudinal Study of 2009 (HSLS: 2009; n = 21,444) was used. The analytic sample was 1,757 (48% female, 52% male) Black (13.8%), Hispanic (26.7%), and White (59.6%) students who scored in the top 10% of their race group on the mathematics achievement test. Hierarchical logistic regression models were developed for each race/ethnicity group to examine the relationships of demographic and expectancy-value variables with STEM persistence status. Science attainment value, science intrinsic value, and STEM utility value were predictive of STEM persistence, but these variables operated differently in groups of Black, Hispanic, and White students. Implications for educators include the need for ways to improve perceptions of science identity and awareness of the utility of science and mathematics courses. © 2013 Wiley Periodicals, Inc. *Sci Ed* **98:**216–242, 2014

Correspondence to: Lori Andersen, Department of Curriculum and Instruction, Kansas State University, Manhattan, KS 66506, USA; e-mail: landersen@ksu.edu

INTRODUCTION

Recent reports have documented an urgent need for science, technology, engineering, and mathematics (STEM) innovators and experts in the United States (National Academy of Sciences, 2007; National Science Board, 2010). However, a much smaller proportion of U.S. students major in the sciences or engineering compared to other countries, and 35% of the PhDs in the U.S. STEM workforce are foreign-born (Atkinson & Mayo, 2011). The acute underrepresentation of minorities in these disciplines is evidence of a large amount of undeveloped talent in these populations. In 2008, Blacks and Hispanics were underrepresented by more than 50% in undergraduate engineering programs compared to their representation in the 18–24-year-old U.S. population, whereas White students are overrepresented by more than 10% (National Science Foundation, 2012). These same levels of underrepresentation also exist in gifted education (U.S. Department of Education, 2008). This disproportionate representation is evidence that the potentials of Black and Hispanic students who have high ability are not being developed.

Future scientists, mathematicians. and engineers should come from the talent pool consisting of *all* students who have high ability or demonstrate superior performance in mathematics and science. Demographic trends in the United States indicate that population diversity is rapidly increasing. Understanding the variables that facilitate STEM persistence for talented Black and Hispanic students is important, not only to provide equitable outcomes for these students compared to the outcomes attained by their White and Asian peers but also to ensure the viability of the U.S. STEM workforce. Students must take appropriate science and mathematics coursework in high school to ensure their readiness to enter postsecondary STEM programs (Lynch, 2011). To increase the numbers of high-ability, underrepresented minority (URM) students who enter trajectories of STEM talent development, the process by which these students plan to take the requisite preparatory coursework must be understood. This article presents the results of a study of the variables that predict ninth-grade, high-ability students' STEM persistence plans.

Framework

The Eccles et al. (1983) expectancy value model of achievement-related choices is the theoretical framework for this study. According to this model, students' decisions to persist in taking mathematics and science coursework are determined by their personal assessments of the likelihood of success in, and the relative value that they assign to, the options perceived to be available. Expectations for success in science and mathematics are represented by science and mathematics self-efficacy. Relative importance is described by *subjective task value* (STV) that construes the value of mathematics and science courses in terms of four dimensions: (1) the utility value as related to the student's future goals, (2) the intrinsic value based on enjoyment, (3) the attainment value based on consistency with student identity, and (4) the cost determined by perceptions of time taken away from other activities or the potential negative responses of peers (Eccles, 2009). STV is synthesized based on inputs from culture, socializers, and the individual's experiences. In other words, STV is constructed during the identity-formation process by which adolescents select activities that reflect the salient characteristics of groups with which they identify (Eccles, 2009).

The plans of high-ability, URM, ninth-grade students to continue their studies of mathematics and science were studied because previous research has shown that reentry into the STEM pipeline is rare after high school and that career plans made in high school predict future completion of STEM degrees (Maltese & Tai, 2011; Syed, Azmitia, & Cooper,

2011; Tai, Liu, Maltese, & Fan, 2006). Before high school, all students are in the science and mathematics pipeline by default. In high school, students follow coursework and career preparation paths that were selected based on perceived ability, motivation, and opportunity. A key to postsecondary STEM talent development is appropriate preparatory coursework in high school (Lynch, 2011). Therefore, a better understanding of the variables that affect these selections could facilitate increases in the numbers of URM students who plan to persist.

STEM Persistence Studies

While previous studies have examined variables associated with STEM persistence using national data, attention has generally been focused on the relative deficits of those students who exit the pipeline. The external validity of these studies is limited by the lack of diversity among the participants. By treating race as one of many predictor variables in a model, most researchers have assumed that variables operate identically across all racial and socioeconomic groups (e.g., Maltese & Tai, 2011; Mau, 2003). Extensive reviews of this literature already exist (Lee & Luykx, 2006; Maltese & Tai, 2011). In summary, previous research has identified deficits in preparatory coursework as a reason why students exit the STEM pipeline (Lee & Luykx, 2006). Early interest was identified as a predictor of who earned a STEM degree (Tai et al., 2006). Taking a greater number of, and more rigorous, mathematics and science courses increased the chances of pursuing a STEM degree (Maltese & Tai, 2011). Fewer Black and Hispanic students completed advanced coursework in mathematics and science compared to their Asian and White peers. However, those who did were equally as likely to complete STEM degrees (Tyson, Lee, Borman, & Hanson, 2007). Students from underrepresented groups have been shown to be at greater risk of leaving a STEM major (Bonous-Harnmarth, 2000). Thus, previous research has revealed the required academic paths (advanced high school mathematics and science) to achieve and the demographics of who was more likely to achieve a STEM degree (Asian, White, and higher socioeconomic status [SES]), but has not examined why many URM high school students who have high ability in mathematics and science take these courses or pursue these degrees. One group of researchers found that career considerations preceded course-taking plans for Black high school students. This finding places the causal order of career choice and course taking asserted by previous research in question (Lewis & Connell, 2005; Thompson & Lewis, 2005). Nonetheless, previous research has not separated the variables that influence persistence by race, thus separate group analyses are necessary to understand and compare how predictor variables operate in different groups (Lee & Luykx, 2006). This study aims to fill this gap in the literature.

Expectations for Success. Persistence is predicted by students' expectations for success in STEM. These expectations were often operationalized as *domain-specific self-efficacy*, or confidence in the ability to successfully complete tasks within a domain. Self-efficacy was more important than achievement to occupational choice decisions (Bandura, Barbaranelli, Caprara, & Pastorelli, 2001; Eccles, 2005). Students who had higher self-efficacy or an interest in mathematics and science were more likely to continue studies of those subjects, after controlling for achievement and SES (Simpkins, Davis-Kean, & Eccles, 2006). Mathematics self-efficacy and academic proficiency of eighth-grade students predicted who would persist in aspiring to a science and engineering career (Mau, 2003). However, the participants in these studies were predominantly White and of mixed ability. In large samples of middle school students, mathematics and science self-efficacy was related to goals and

intentions for Mexican American, eighth-grade students (Navarro, Flores, & Worthington, 2007) and for inner-city, low-SES students (Fouad & Smith, 1996). In summary, previous research supports the importance of self-efficacy to occupational choice and course-taking plans in groups of mixed-ability students. But little is known about the relative importance of domain-specific self-efficacy to high-ability students' persistence plans.

Subjective Task Value. Two studies have been conducted using data from the National Education Longitudinal Study of 1988 (NELS: 88) that examined the effects of STV on persistence. First, early interest in a STEM career was sufficient to sustain students in the pipeline. Students who planned on pursuing a STEM career were more than twice as likely to earn a college degree in the sciences than students who did not have such plans, after controlling for student background and mathematics achievement (Tai et al., 2006). Eighth-grade students' perceptions of the science utility value, a component of STV, was a better predictor of who would complete a STEM degree than mathematics or science achievement test scores (Maltese & Tai, 2011). These studies support the predictive value of the intrinsic and utility value components of STV. However, no previous studies were found that examined the predictors of STEM persistence within a nationally representative sample of high-ability students.

Few studies have examined racial or ethnic differences in STV. Zarrett and Malanchuk (2005) studied Black students' decisions to pursue careers in information technology. Black students were equally as likely to consider a career in computers as White students. Students' perceived ability, value of a domain, and the influence of socializers and peers on students' decision to pursue an information technology career were significant effects. These findings support the relevance of STV to Black students' career decisions.

There have been no empirical studies of high-ability high school students' STV for STEM. According to expectancy-value theory, students who place a high STV on mathematics and science should be motivated to take such coursework. STV will vary within and across racial and ethnic groups because of the differential effects of culture and socializers on student identities (Eccles, 2009; Simpkins & Davis-Kean, 2005). For example, the compatibility of doing mathematics and science with the individual's identity is the source of attainment value; therefore, components of that identity such as race, ethnicity, gender, and culture will affect the STV that is constructed for science and mathematics.

Race, Ethnicity, Culture, and STV. The four components of STV are each affected by the racial, ethnic, and cultural identity of the student and the interactions of these attributes with STEM culture. For example, a lack of same-race role models or prominent historical figures in science or mathematics may prevent minority students from identifying with STEM domains. These students may feel as though they must be assimilated and give up their racial identity to succeed (Cooper, 2011). Many minority students may be less likely to view science and mathematics coursework as having a high utility value because of a lack of evidence of the successes of people like themselves, as compared to White male students who are presented with ample evidence of the successes of similar people in science (Hines, 2003). Science and mathematics careers may not seem like reasonable possibilities for personal goals to minority students (Archer et al., 2010; Archer, Hollingworth, & Halsall, 2007). Lewis and Connell (2005) found that a majority of Black students' science and mathematics course-taking decisions were based on utility value or interest. Lower utility values caused by a lack of connection between STEM courses and students' personal goals contribute to a lower STV and reduce the likelihood of plans to persist.

Incompatible Identities. Adolescence is a period focused on identity formation, including the development of academic and occupational identities (Erikson, 1968). Students develop a better sense of their relative competencies and the values that self-esteem is based on during this process (Wigfield & Wagner, 2005). Occupations are an important source of identity and adolescents choose future occupations based, in part, on how well their perceptions of who typically performs that kind of work, and what that work entails, fit with their identities (Bandura et al., 2001). Science is a subculture of White, male, Western culture (Barba, 1998; Hines, 2003). Stereotypes that are associated with STEM are likely to conflict with components of students' gender, ethnic, or racial identities and prevent the integration of science into their identities (Archer et al., 2007, 2010; Taconis & Kessels, 2009). For example, the culture of science is perceived to be masculine, competitive, individualistic, cutthroat, and isolated whereas many minority students' learning styles demonstrate preferences for collaboration, group work, cooperation, and social learning (Ford, 2011; Heilbronner, 2011; Seymour & Hewitt, 1997). Furthermore, STEM is often associated with social attributes that are undesirable to adolescents, which discourages the selection of such occupations. These points of potential cultural conflict mean that minority students may have lower degrees of identification with, and thus a lower degree of attainment value for, science than nonminority students. Attainment value and STV are reduced when science identity is lower, which inhibits persistence. Thus, differences in the STV that students construct for science and mathematics may explain differences in persistence plans.

This study investigated the expectations for success and the STV that high-ability students have for science and mathematics by comparing the effects of factors such as self-efficacy, attainment value, utility value, intrinsic interest, and cost on these students' plans to persist. The STEM persistence plans of high-ability students were hypothesized to be a function of these variables. Based on the Eccles et al. (1983) model, it was hypothesized that students who have high expectations for success, have intrinsic interest, see a high degree of utility in taking science and mathematics courses related to their future goals, find science and mathematics consistent with their identity, and have positive perceptions of the cost of taking science and mathematics courses are more likely to plan to persist. The current investigation explores the relative importance of these factors.

Research Questions

This investigation used a sample of high-ability, ninth-grade students to study variables that may be associated with their plans to persist in STEM. Based on the Eccles et al. model and the review of the literature, the following hypotheses were made:

1. Each of the two measures of individuals' expectations for success in STEM, mathematics and science self-efficacy, will be significantly and positively related to persistence plans after controlling for SES, gender, and mathematics achievement.
2. Each of the five measures of STV—STEM utility value, mathematics and science intrinsic values, and mathematics and science attainment values—will be significantly and positively associated with persistence plans after controlling for SES, gender, and mathematics achievement.
3. A positive perception of the cost of taking mathematics and science courses will be significantly and positively associated with persistence plans after controlling for SES, gender, and mathematics achievement.

METHODOLOGY

Sample

The High School Longitudinal Study of 2009 (HSLS: 2009; Ingels et al., 2011) is a secondary longitudinal study from the National Center for Education Statistics (NCES). These data came from the base year of HSLS: 2009. The sample was representative of ninth-grade students in public and private schools in the United States in 2009. Within each of the 944 participating schools, a stratified random sample of students was selected based on race/ethnicity. An average of 27 students per school were selected, and the total number of students who participated in the study was 21,444. Data were collected during the fall of the ninth grade. For this study, the analytic sample was reduced to the group of Black, Hispanic, and White students who were identified as having high ability in mathematics or science. The group consisted of 1,757 students (13.8% Black, 26.7% Hispanic, and 59.6% White) of whom 48.5% were female and 59.6% were male. Each group was analyzed separately.[1]

Missing Data

A total of 23 variables from HSLS: 2009 were used. Missing data percentages on items ranged from 0% to 4.8%, with a mean of 2.4% (SD = 1.2%). The mechanism for missing data was assumed to be missing at random (Enders, 2010). Missing values for the independent variables were replaced using the expectation maximization (EM) procedure in SPSS 20.

Weights

The analyses were based on weighted samples that were created to adjust for oversampling bias and nonresponse (NCES, 2011). The first-year student weight (W1student) was used. To compensate for the way that SPSS calculates standard errors for weighted data based on population size rather than sample size, the weight was normalized and divided by the design effect (NCES, 2011).

Variables

Grouping Variables. The analytic sample was selected using the variables of race and high-ability status. Race was provided by NCES, and high-ability status was operationalized as students who scored in the top 10% of their race group on the mathematics achievement test. This threshold was selected based the recent definition of giftedness as performance in the top 10% of the peer group (NAGC, 2011). Group-specific norms are recommended for the identification of ability in underrepresented groups (e.g., Lohman, 2005). Students who met the mathematics achievement test criterion were identified as high ability (Table 1). The analytic sample was reduced to the 1,757 students who met the high-ability criteria.

Independent Variables. Eleven independent variables were used to create a model for STEM persistence. Six of these variables were provided by NCES, and four others were created by the researchers. The development of each scale is described in this section.

[1] An analysis of the entire group that included interactions of each variable with race revealed no significant interactions due to a lack of sufficient sample size to support a logistic regression analysis with a large number of predictor variables. Race has three levels; therefore, adding the interactions of 11 variables with race created 22 additional independent variables.

TABLE 1
High Ability Criteria by Race

Variable	White	Black	Hispanic
Mathematics achievement score	55.98	49.59	51.56

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Values not weighted.

Socioeconomic Status. A standardized, continuous, composite variable was created by NCES based on parent/guardian education, occupation, and family income. Data for nonresponding parent/guardians were imputed by NCES.

NCES-Created Scales. Certain groups of items in the student survey were designed by NCES to be used as psychological scales (Ingels et al., 2011). The Eccles et al. (1983) expectancy-value framework was used in the design of HSLS: 2009. Therefore, these scales were used in the present study. These scales included: mathematics self-efficacy, science self-efficacy, mathematics identity, and science identity (see Appendix A). All questionnaire items were reverse coded such that larger scale values corresponded to positive attributes (Ingels et al., 2011). The reliability of each scale was assessed using Cronbach's alpha; scales were required to meet a minimum threshold value of .65. Scales were created and then standardized to a mean of zero and standard deviation of 1.0. These scales were created by NCES and used by the researchers for the present study. A summary of all scales and reliability coefficients is presented in Appendix B.

Mathematics and Science Self-Efficacy. Two scale scores represented mathematics and science self-efficacy, respectively. The items used to construct this scale asked students about their beliefs in their abilities to be successful in the current mathematics and science course. The mathematics and science self-efficacy scales had Cronbach's alphas of .90 and .88, respectively (Ingels et al., 2011).

Attainment Value. Attainment value is based on the consistency of a mathematics or science identity with the student's identity, thus the mathematics and science identity scales that were created by NCES were used to represent mathematics and science attainment value, respectively. Students were asked how well they agreed with statements such as "You see yourself as a math (science) person" and "Others see you as a math (science) person." Mathematics attainment value had a reliability of .84, and science attainment value had a reliability of .83 (Ingels et al., 2011).

Researcher-Created Scales.

Utility and Intrinsic Value. The researchers constructed scale scores for utility and intrinsic value. Student responses to a series of questions that probed the reasons why students planned to take more mathematics or science courses during high school were used to construct scales for the utility and intrinsic value of mathematics and science courses. Eight of these reasons were identified as representative of utility value or intrinsic value based on item content analysis (Table 2). Principal components analysis was used with the set of eight items for dimension reduction, and three standardized factor scores were created that were labeled STEM utility value, mathematics intrinsic value, and science intrinsic value (Table 2).

TABLE 2
Summary of Items and Factor Loadings for Varimax Orthogonal Three-Factor Solution for Utility and Intrinsic Value Items ($N = 19{,}259$)

Item	Factor Loading 1	Factor Loading 2	Factor Loading 3	Communality
F02I Plans to take more mathematics courses because it will help to get into college.	**.83**	.06	.18	.72
F02J Plans to take more mathematics courses because it will be useful in college.	**.81**	.09	.23	.72
F05I Plans to take more science courses because it will help to get into college.	**.84**	.21	.03	.74
F05J Plans to take more science courses because it will be useful in college.	**.83**	.25	.05	.76
F05H Plans to take more science courses because he/she enjoys studying science.	.15	**.87**	.12	.80
F05E Plans to take more science courses because he/she is good at science	.21	**.84**	.16	.77
F02H Plans to take more mathematics courses because he/she enjoys studying mathematics.	.08	.15	**.86**	.76
F02E Plans to take more mathematics courses because he/she is good at mathematics.	.20	.11	**.84**	.76
Eigen value	3.63	1.33	1.06	
Percentage of variance	45.40	16.66	13.30	

Note: A value in bold indicates the highest factor loading.

STEM Utility Value. Four of the eight questions asked students whether they planned to take future mathematics or science courses because they needed the courses to get into college or because the courses were useful for college. These four items loaded on one factor (Table 2). These factor loadings were used to create a standardized scale score for STEM utility value. The reliability for this scale was .87.

Intrinsic Value. Four of the eight questions asked students whether they planned to take future mathematics and science courses because they enjoyed or were good at mathematics or science. The two science items loaded on factor two and the two mathematics items loaded on factor three. The factor loadings were used to create a standardized scale score for mathematics intrinsic value, and the two science variables were used to create a scale score for science intrinsic value (Table 2). The two scales had reliabilities of .68 and .73, respectively.

Cost. The researchers also constructed a scale for cost. Four questions concerned the impact of spending a lot of time and effort in mathematics and science classes on the amount of time available to spend with friends, time to spend on other activities, popularity, and being made fun of. The four items were reverse coded such that higher values corresponded to more positive perceptions of cost and were used to create the cost scale. The Cronbach's alpha for this scale was .75; the scale was normalized to a mean of zero and a standard deviation of 1.0.

Dependent Variable. The dependent variable of this study was a dichotomous variable that indicated STEM pipeline status. Students who identified the occupation they expected to have at age 30 as (1) computer and mathematical; (2) architecture and engineering; (3) life, physical, and social sciences; or (4) healthcare practitioners and technical occupations were identified as having planned to persist. An alternate criterion for selection was devised because a large number of students (28.2%) responded with "don't know." If a student planned on taking 4 years of mathematics, 4 years of science, and at least one Advanced Placement or International Baccalaureate mathematics or science course during high school, the student was included. Students who met either of the two criteria—identification of a future STEM occupation or indication of intent to a plan to persist—were assigned the dependent variable value of "planned to persist."

Logistic Regression Model

The goal was to investigate the role that expectations for success and STV had on student persistence plans within each group. A logistic regression examines the effects of the many independent variables on one dichotomous, dependent variable (Hosmer & Lemeshow, 2000). STEM persistence status was the dependent variable. Each regression was performed in steps with SES entered in the first step, gender in the second step, and mathematics achievement test score in the third step. The group of expectancy value variables was entered in the fourth step. The variables were entered stepwise to retain only significant predictors in the model at each step. This allowed the examination of how the relationships between significant variables and the dependent variable evolved as additional factors were added.

The decision was made to separate the sample by race/ethnicity group and perform separate logistic regression analysis because the power of the analysis was limited. The number of independent variables in the model was so large that the introduction of interaction variables for each of the three levels of race/ethnicity with the 11 predictor variables created 22 potential interaction variables. The sample size, though considerable, was insufficient to support the simultaneous testing of all interaction variables. Therefore, separate analyses were conducted for each level of race/ethnicity to explore potential differences in the operation of the expectancy value model. Although this method fails to provide tests of statistical significance regarding differences in the regression coefficients or odds ratios (OR) between groups, it does provide a starting point for further investigations into between group differences. An implication of this methodological choice is that between-group differences should be considered tentatively and further analyses are needed.

Validity

Threats to Internal Validity. This study had several threats to internal validity. First, although the researchers took care to select the survey items that best reflected the constructs within the expectancy value model, these items were all worded to describe students' expectancies and values about the mathematics and science courses that they were taking in 2009 and may not reflect their values about these subjects in other contexts, such as real-world applications. Second, a lack of a standardized measure of science achievement that led to the use of other variables as a proxy for science achievement. Third, the occupation classification method available in the HSLS: 2009 public use database limited the researchers' ability to precisely sort occupations into STEM and non-STEM categories. Fourth, manual adjustments were made to values calculated by SPSS 20 using the procedures recommended by NCES (Ingels et al., 2011) because the complex study design

and weighting used in this data set affected statistical significance measures. Next, the method used to handle missing values was limited by the capabilities of SPSS. EM was used instead of multiple imputations. Finally, the model created in this analysis is only one possible model of STEM persistence plans, many other models are possible and another model may better explain the variations in persistence.

Threats to External Validity. The operationalization of high ability is a threat to external validity. The design of HSLS: 2009 determined what information was available to identify students in the sample as having high ability in mathematics and science. This operationalization may differ from other definitions and thus impacts the results.

RESULTS

Students were identified as having high ability as described in the Methods section. Using multiple criteria for identification acknowledged findings in the literature regarding the importance of domain-specific criteria and group-based norms for identification of high ability (e.g., Lohman, 2005). The criterion for identification was different for each race group (Table 1).

The goal of this study was to identify the significant predictors of plans to persist for ninth-grade, high-ability students for each race/ethnicity group. Descriptive statistics for the predictor variables by persistence plan status and overall are displayed for each group (Tables 3–6). Examination of these data revealed differences between the three high-ability groups. In the Black group, persisters scored significantly higher than nonpersisters in mathematics achievement, science intrinsic value, and science attainment value. In the Hispanic group, persisters scored significantly higher than nonpersisters in STEM utility value and science attainment value. In the White group, there were significant differences between persisters and nonpersisters on science self-efficacy, science intrinsic value, mathematics attainment value, and science attainment value. All differences favored the persister group.

In Table 6, the means for each race/ethnicity group are compared. SES and science attainment value evidence large differences between White students and Black or Hispanic students. The Black group and the Hispanic group had similar scores on some variables such as mathematics self-efficacy, mathematics intrinsic value, cost, and mathematics attainment value, but these groups differed more on the science-related variables. The Hispanic group was more similar to the White group than the Black group in terms of science-related variables. Importantly, the selection of these high-ability students based on mathematics achievement test scores at the 90th percentile or higher did not produce range restriction in the self-efficacy variables; the descriptive statistics do not indicate range restriction that would attenuate correlations.

Bivariate correlations were calculated for each pair of continuous predictor variables within each group (Tables 7–9). None of the sizes of the correlation coefficients raised concerns about collinearity (maximum correlation = .60). The mathematics-related variables were moderately correlated, and the science-related variables (self-efficacy, intrinsic value, and attainment value) were moderately correlated.

Hierarchical (stepwise) logistic regression analyses were used to examine demographic variables (SES, gender, and mathematics achievement) that previous research has identified as predictive of STEM persistence. In the third step, the expectancy-value factors were added (Tables 10–15). The regressions were run stepwise backward using the Wald criterion, and the resulting models were verified using stepwise forward methods which confirmed the results.

TABLE 3
Descriptive Statistics for Predictor Variables as a Function of STEM Pipeline Status for High-Ability Black Students ($n = 221$)

Variable	Persisters ($n = 119$) M (SE)	Nonpersisters ($n = 102$) M (SE)	Overall ($n = 221$)	$\chi^2(1)$ or $t(349)$	p
Female[a]	82	41	123	5.551	.019
Male[a]	50	48	98		
SES[b]	0.22 (0.10)	−0.04 (0.06)	0.10 (0.06)	1.346	.184
Mathematics achievement	55.04 (0.31)	53.12 (0.27)	54.15 (0.26)	2.088	.042
Mathematics self-efficacy[b]	0.58 (0.10)	0.32 (0.07)	0.46 (0.07)	1.131	.263
Science self-efficacy[b]	0.38 (0.18)	0.02 (0.06)	0.21 (0.10)	1.223	.227
STEM utility value[b]	0.53 (0.10)	0.35 (0.07)	0.44 (0.06)	0.712	.480
Science intrinsic value[b]	0.10 (0.16)	−0.44 (0.14)	−.15 (0.10)	2.019	.049
Mathematics intrinsic value[b]	0.33 (0.13)	0.23 (0.08)	0.28 (0.08)	0.280	.781
Cost[b]	0.00 (0.21)	0.30 (0.06)	0.14 (0.12)	1.040	.303
Mathematics attainment value[b]	0.59 (0.09)	0.47 (0.06)	0.53 (0.06)	0.522	.604
Science attainment value[b]	0.46 (0.11)	−0.20 (0.09)	0.15 (0.07)	2.625	.011

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
[a]Frequency.
[b]Standardized score with an approximate mean of zero and approximate standard deviation of one.

Overall Model

SES, Gender, and Mathematics Achievement. The direct effect of SES on ninth-grade high-ability students' plans to persist in STEM was examined. SES did not significantly predict planned STEM persistence for any group of high-ability students; students from higher SES households were not significantly more likely to plan to persist. Therefore, this variable was not retained in subsequent models. The effect of gender on persistence plans was not statistically significant for any group of high-ability students, and it was not retained in subsequent models. Mathematics achievement did not significantly predict persistence for Hispanic or White students, but was a significant predictor for Black students. However, the selection of students using mathematics achievement as a criterion resulted in a restricted range for this variable; thus these effects were most likely attenuated for all groups.

Expectancy-Value Variables. The individual expectations of success variables, science and mathematics self-efficacy, were not significant predictors of persistence plans for any

TABLE 4
Descriptive Statistics for Predictor Variables as a Function of STEM Pipeline Status for High-Ability Hispanic Students ($n = 351$)

Variable	Persisters ($n = 217$) M (SE)	Non-persisters ($n = 134$) M (SE)	Overall ($n = 351$)	$\chi^2(1)$ or $t(219)$	p
Female[a]	103	77	180	3.022	.082
Male[a]	82	89	171		
SES[b]	0.01 (0.06)	−0.19 (0.06)	−0.08 (0.05)	1.393	.167
Mathematics achievement	56.15 (0.25)	55.02 (0.31)	55.62 (0.19)	1.611	.110
Mathematics self-efficacy[b]	0.48 (0.05)	0.54 (0.16)	0.51 (0.08)	0.333	.740
Science self-efficacy[b]	0.50 (0.09)	0.24 (0.10)	0.37 (0.07)	1.500	.137
STEM utility value[b]	0.59 (0.07)	0.02 (0.10)	0.32 (0.06)	3.327	.001
Science intrinsic value[b]	0.50 (0.14)	0.13 (0.13)	0.32 (0.09)	1.603	.112
Mathematics intrinsic value[b]	0.32 (0.12)	0.23 (0.16)	0.28 (0.10)	0.396	.693
Cost[b]	0.14 (0.09)	0.10 (0.05)	0.12 (0.05)	0.229	.819
Mathematics attainment value[b]	0.68 (0.06)	0.54 (0.07)	0.61 (0.05)	0.856	.394
Science attainment value[b]	0.64 (0.11)	0.12 (0.16)	0.39 (0.09)	2.879	.005

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
[a]Frequency.
[b]Standardized score with an approximate mean of zero and approximate standard deviation of one.

high-ability group. Of the group of STV variables, science attainment value was significant for all three groups, whereas STEM utility value was identified as significant predictors of persistence only for Hispanic students. Students who held a higher attainment value for science were more likely to plan to persist. The degree to which students identified with science was predictive of plans to persist. No other variables were significant predictors of persistence.

The pseudo R^2 for the final models were .271, .195, and .178 for the Black, Hispanic, and White groups, respectively.

DISCUSSION

The complex study design of HSLS: 2009 allows for inferences to be made to the larger population of U.S. students who were in the ninth grade in Fall 2009. The 1,757 students represent 346,096 high-ability ninth graders in 2009. This is the group to which inferences are made.

TABLE 5
Descriptive Statistics for Predictor Variables as a Function of STEM Pipeline Status for High-Ability White Students ($n = 1,185$)

Variable	Persisters ($n = 804$) M (SE)	Nonpersisters ($n = 381$) M (SE)	Overall ($n = 1,185$)	$\chi^2(1)$ or $t(1183)$	p
Female[a]	366	180	546	0.061	0.805
Male[a]	424	215	639		
SES[b]	0.73 (0.03)	0.66 (0.04)	0.71 (0.02)	0.680	.497
Mathematics achievement	60.50 (0.13)	59.91 (0.19)	60.30 (.12)	1.235	.218
Mathematics self-efficacy[b]	0.68 (0.03)	0.47 (0.05)	0.61 (0.03)	1.701	.090
Science self-efficacy[b]	0.68 (0.03)	0.21 (0.06)	0.52 (0.03)	3.716	.000
STEM utility value[b]	0.57 (0.03)	0.38 (0.04)	0.51 (0.02)	1.692	.092
Science intrinsic value[b]	0.64 (0.04)	−0.07 (0.05)	0.40 (0.03)	4.590	.000
Mathematics intrinsic value[b]	0.76 (0.04)	0.56 (0.05)	0.69 (0.03)	1.242	.216
Cost[b]	0.35 (0.04)	0.12 (0.05)	0.27 (0.02)	1.783	.076
Mathematics attainment value[b]	0.95 (0.03)	0.60 (0.04)	0.83 (0.02)	3.043	.003
Science attainment value[b]	0.82 (0.03)	0.08 (0.06)	0.57 (0.03)	5.819	.000

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
[a]Frequency.
[b]Standardized score with an approximate mean of zero and approximate standard deviation of one.

Research Hypotheses

The goal of this study was to examine the dynamic processes by which ninth-grade, high-ability students made STEM persistence plans within each race/ethnicity group. It was hypothesized that expectations for success, STV, and cost would be significantly and positively related to persistence. The results of this analysis partially support the hypotheses. Hypothesis 1 was not supported in the final model. Neither of the self-efficacies predicted persistence plans. Hypothesis 2 was partially supported. The final model showed that three components of STV were positively and significantly related to persistence in the final models. One significant predictor was common to the three groups; science attainment value was a significant predictor of persistence plans for Black, Hispanic, and White students. STEM utility was a significant predictor for Hispanic students, but not for Black or White students. Science interest value and mathematics attainment value were retained in the model for White students but had p values of .097 and .088, respectively. Hypothesis 3 was not supported. The cost variable was not a significant predictor of persistence for any group.

TABLE [redacted]
Comparison of Ninth-Grade, High-Ability Students Across Race/Ethnicity Groups

Variable	White ($n = 1{,}185$) M (SD)	Black ($n = 221$) M (SD)	Hispanic ($n = 351$) M (SD)
Female[a]	546	123	180
Male[a]	639	98	171
SES[b]	0.71 (0.71)	0.10 (0.72)	−0.08 (0.74)
Mathematics achievement	60.30 (3.41)	54.15 (3.43)	55.62 (3.54)
Mathematics self-efficacy[b]	0.61 (0.88)	0.46 (0.84)	0.51 (0.97)
Science self-efficacy[b]	0.52 (0.93)	0.21 (1.06)	0.37 (0.88)
STEM utility value[b]	0.51 (0.82)	0.44 (0.93)	0.32 (0.91)
Science intrinsic value[b]	0.40 (1.19)	−0.15 (1.03)	0.32 (1.19)
Mathematics intrinsic value[b]	0.69 (1.11)	0.28 (1.19)	0.28 (1.18)
Cost[b]	0.27 (0.91)	0.14 (1.04)	0.12 (0.86)
Mathematics attainment value[b]	0.83 (0.82)	0.53 (0.84)	0.61 (0.85)
Science attainment value[b]	0.57 (0.96)	0.15 (0.96)	0.39 (0.95)

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
[a]Frequency.
[b]Standardized score with an approximate mean of zero and approximate standard deviation of one.

These findings suggest that ninth-grade, high-ability students who have a higher attainment value for science are more likely to plan to persist in STEM (OR of 2.479, 1.719, and 1.898 for Black, Hispanic, and White students, respectively). For Hispanic students, a higher utility value was a predictor of persistence (OR = 1.95), whereas for Black students a higher mathematics achievement was a predictor (OR = 1.254). Mathematics and science self-efficacy did not play a significant role in persistence plans for these students. This finding contradicts other research that supported mathematics self-efficacy as predictive of STEM persistence in mixed-ability groups of students (e.g., Mau, 2003; Simpkins et al., 2006). However, no previous studies have examined such effects in groups of high-ability students.

Effect of SES and Gender

SES was not a significant predictor of persistence for high-ability students within each race/ethnicity group. This finding is encouraging because it implies that low-SES students are not less likely to persist. However, the descriptive statistics for each group show a large disparity in SES between the groups. The mean SES for high-ability White students was 0.71, whereas the mean SES for high-ability Black or Hispanic students were 0.10 and –0.08, respectively. Furthermore, the overall persistence rate of White (67%) student was substantially larger than for Black (53%) or Hispanic (53%) students. Thus, an analysis of the overall group would show that SES is correlated to persistence because of the effect of the White group.

Another interesting finding of this study is that gender was not a significant predictor of persistence plans for any group. This suggests that among high-ability students there is

TABLE 7
Intercorrelations for Predictor Variables of Planned STEM Persistence (High-Ability, Black Students)

Measure	1	2	3	4	5	6	7	8	9	10
1. SES	1									
2. Mathematics self-efficacy	−.04	1								
3. Science self-efficacy	.17	.17	1							
4. STEM utility value	.09	−.08	.15	1						
5. Science intrinsic value	−.07	−.06	.29*	−.02	1					
6. Mathematics intrinsic value	.03	.41**	−.23	−.32*	.00	1				
7. Cost	−.09	−.03	.21	.08	.10	−.15	1			
8. Mathematics attainment value	−.10	.60**	.10	−.15	.00	.51**	.09	1		
9. Science attainment value	.13	.15	.53**	.19	.42**	−.04	.08	.27	1	
10. Mathematics achievement score	.11	.05	.01	.11	−.08	−.13	−.14	−.01	−.01	1

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
*$p < .05$, **$p < .01$, ***$p < .001$.

no evidence of gender stereotyping with regard to STEM persistence plans. However, this result may be affected by the inclusion of the life sciences and health sciences in the STEM category because these domains tend to be pursued by larger numbers of females. The chi-square test of the effect of gender on persistence in the Black group showed that gender was significantly related to persistence ($\chi^2(1) = 5.551$, $p = .019$), but the effect was not significant in the logistic regression analysis, which may indicate a lack of power. In contrast to this finding for ninth-grade students, the persistent underrepresentation of women in the STEM fields implies that females' expectancies and values for STEM may change after the ninth grade and negatively impact their persistence plans. This is supported by the findings of Archer and her colleagues (2007, 2010), who cite disparities between cultural expectations of femininity and stereotypical images of scientists as barriers to female participation in STEM. They found these effects earlier than the ninth grade, although they did not study high-ability students.

Effect of Mathematics Achievement

Students were selected as having high ability if the mathematics achievement test score was at the 90th percentile or above for their race/ethnicity group. The mathematics achievement score was included to control for differences in persistence due to mathematics ability.

TABLE 8
Intercorrelations for Predictor Variables of Planned STEM Persistence (High-Ability, Hispanic Students)

Measure	1	2	3	4	5	6	7	8	9	10
1. SES	1									
2. Mathematics self-efficacy	−.18	1								
3. Science self-efficacy	.08	.31**	1							
4. STEM utility value	.24*	−.23*	.04	1						
5. Science intrinsic value	−.02	.02	.42**	.00	1					
6. Mathematics intrinsic value	.11	.28**	.05	−.07	−.02	1				
7. Cost	.21*	.18*	.24*	.09	.07	.15	1			
8. Mathematics attainment value	.00	.39**	.18	−.16	.11	.53**	.06	1		
9. Science attainment value	.19	−.04	.39**	.18	.45**	−.04	.12	.16	1	
10. Mathematics achievement	.24*	.18	.15	.04	.16	.32**	.12	.35**	.12	1

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

The effect of mathematics achievement was only significant for high-ability Black students. The OR of 1.254 indicates that a one-point increase on the mathematics achievement test equated to a 25.4% greater chance of the student planning to persist. This test had a 70-point maximum, and the mean score for high-ability, Black students was 54.15 points. This result may reflect the fact that the identification criterion for giftedness used by schools depends on global norms and not on group-based norms. Thus, the Black students who were in the 90th percentile or greater for their own group would not have been identified as having high ability in their own schools because the 90th percentile cutoff score for the White group was substantially larger. For example, in this study the mean score for the high-ability Black group (54.15) was below the 90th percentile score for Whites (55.98). Thus, only the Black students who have the highest scores for their group will reach identification thresholds in a school with a large White majority. Such students would not be identified as having high ability and may not self-identify as high-ability students.

Self-Efficacy and Persistence Plans

One explanation as to why mathematics and science self-efficacy were not significant predictors of persistence plans in this sample may be because the self-efficacy measures were specific to students' perceptions of their ability to succeed in their ninth-grade coursework. Self-efficacy regarding school science and mathematics may have a weak relationship

TABLE 9
Intercorrelations for Predictor Variables of Planned STEM Persistence (High-Ability, White students)

Measure	1	2	3	4	5	6	7	8	9	10
1. SES	1									
2. Mathematics self-efficacy	.01	1								
3. Science self-efficacy	.02	.38**	1							
4. STEM utility value	−.01	−.01	.08	1						
5. Science intrinsic value	.11	.17*	.43**	−.01	1					
6. Mathematics intrinsic value	.06	.40**	.07	−.08	.09	1				
7. Cost	.01	.22**	.29**	.03	.16*	.14*	1			
8. Mathematics attainment value	−.04	.50**	.25**	.01	.10	.50**	.14*	1		
9. Science attainment value	.07	.17**	.45**	.10	.56**	.04	.19**	.29**	1	
10. Mathematics achievement score	.12	.22**	.18**	−.07	.06	.20**	.07	.27**	.08	1

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
*$p < .05$, **$p < .01$, ***$p < .001$.

TABLE 10
Nested Models for the Planned STEM Persistence of High-Ability, Black Ninth-Grade Students ($N = 221$)

	Model			
Variable	1	2	3	4
SES	ns	–	–	–
Gender		ns	–	–
Mathematics achievement			1.220	1.254*
Science attainment				2.479*
χ^2			4.592	11.869
$\Delta\chi^2$			4.592*	7.277**
df			1	2
Δdf			1	1
Pseudo R^2			.112	.271
ΔPseudo R^2			.112*	.159**

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
*** $p < .001$, **$p < .01$, *$p < .05$.

TABLE 11
Nested Models for the Planned STEM Persistence of High-Ability, Hispanic Ninth-Grade Students ($N = 351$)

Variable	Model 1	2	3	4
SES	ns	–	–	–
Gender		ns	–	–
Mathematics achievement			ns	–
STEM utility				1.950**
Science attainment				1.719*
χ^2	–	–	–	16.046
$\Delta\chi^2$		–	–	16.046***
df	–	–	–	2
Δdf		–	–	2
Pseudo R^2	–	–	–	.195
ΔPseudo R^2		–	–	.195***

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

TABLE 12
Nested Models for the Planned STEM Persistence of High-Ability, White Ninth-Grade Students ($N = 1{,}148$)

Variable	Model 1	2	3	4
SES	ns	–	–	–
Gender		ns	–	–
Mathematics achievement score			ns	–
Science attainment				1.898**
χ^2	–	–		31.105
$\Delta\chi^2$		–		31.105***
df	–	–		1
Δdf		–		1
Pseudo R^2	–	–		.178
ΔPseudo R^2		–		.178***

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.
$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

with students' plans to pursue a STEM career. Expectations for success in a career may not be adequately represented by school subject self-efficacies. An alternative explanation is that students do not make connections between school science and mathematics and their future career plans. This explanation is supported by the findings of Archer et al. (2010), who found that school science was viewed by students as completely different than "real" science. Further evidence for this explanation is the relatively low STEM utility value scores

TABLE 13
Logistic Regression Models for STEM Persistence of High-Ability Black Students

Model	Variable	*B*	SE	Wald	Odds Ratio	CI	*p*
1	SES	.545	0.410	1.764	1.725	0.772, 3.856	.184
2	Female	.135	0.567	0.056	1.144	0.376, 3.478	.812
3	Mathematics achievement	.199	0.102	3.822	1.220	0.999, 1.489	.051
4	Mathematics achievement	.226	0.110	4.230	1.254	1.010, 1.556	.040
	Science attainment value	.908	0.370	4.193	2.479	1.201, 5.120	.014

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.

TABLE 14
Logistic Regression Models for STEM Persistence of High-Ability Hispanic Students

Model	Variable	*B*	SE	Wald	Odds Ratio	CI	*p*
1	SES	.381	0.276	1.907	1.464	0.852, 2.514	.167
2	Female	.375	0.399	0.885	1.456	0.666, 3.182	.347
3	Mathematics achievement	.095	0.060	2.496	1.099	0.978, 1.236	.114
4	STEM utility	.668	0.246	7.393	1.950	1.205, 3.157	.007
	Science attainment	.542	0.238	5.183	1.719	1.078, 2.741	.023

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.

TABLE 15
Logistic Regression Models for STEM Persistence of High-Ability White Students

Step	Variable	*B*	SE	Wald	Odds Ratio	CI	*p*
1	SES	.137	0.200	0.465	1.146	0.774, 1.697	.496
2	Female	.029	0.283	0.010	1.029	0.591, 1.792	.919
3	Mathematics achievement	.053	0.043	1.519	1.054	.969, 1.146	.218
4	Science interest	.258	0.155	2.760	1.294	0.955, 1.755	.097
	Mathematics attainment	.326	0.191	2.903	1.385	0.952, 2.015	.088
	Science attainment	.641	0.206	9.707	1.898	1.268, 2.841	.002

Source: High School Longitudinal Study of 2009.
Tabulations by the authors.
Data are weighted by W1Student.

for these high-ability students, which indicated that students did not find science very useful for college or career. However, this finding about the relationship between self-efficacy and persistence contradicts other research on STEM persistence, intentions, and goals that has found mathematics self-efficacy to be predictive of persistence plans (Fouad & Smith, 1996; Mau, 2003; Navarro et al., 2007). Notably, these studies did not include the STV variables that were included in this study and were conducted with mixed-ability groups. Further research should examine the predictive value of self-efficacy on STEM persistence plans.

The present study included only high-ability students. Therefore, it could be posited that this group of students has higher self-efficacy and that a restriction of range of this variable attenuated the relationship between self-efficacy and persistence. However, the data show that the correlations between achievement and mathematics or science self-efficacy are generally not significant (for Black or Hispanic students; Tables 7 and 8) or significant but small ($r = .18$ and $.22$, $p < .01$ for White students; Table 9). Furthermore, no restriction of range was observed in the self-efficacies of high-ability students (Table 6), thus the high-ability students in this sample do not appear to have much higher self-efficacies in mathematics and science than other students. Notably, these studies did not include the STV variables that were included in this study and were conducted with mixed-ability groups. Further research should examine the predictive value of self-efficacy on STEM persistence plans.

Subjective Task Value and Persistence Plans

STV has four components: intrinsic value, utility value, attainment value, and cost. Students' development of each component of STV is affected by sociocultural factors, and subsequent differences in STV may affect STEM persistence plans. In this section, each of these components will be discussed.

Intrinsic Value. According to the Eccles et al. (1983) model, the development of students' intrinsic value of science depends on sociocultural factors. Historically, the traditional image of science is one of a quest for knowledge that is motivated by an intrinsic desire to know, even if the knowledge may not be relevant or useful and the pursuit of knowledge for its own sake may be viewed as a luxury of the privileged (Brickhouse, 1994). In this sample of students with high ability in science and mathematics, it was expected that science and mathematics' intrinsic values would be significantly above average. However, this expectation was not supported by the descriptive statistics for each high-ability group. Significant differences were found in both science intrinsic value favoring White students (Table 6) over Black or Hispanic students and persisters over nonpersisters in both groups (Tables 3–5). Furthermore, the mean science intrinsic scale z score for the Black group was –0.15, which is quite low for a high-ability sample. Thus, high-ability Black students had a much lower sense of science intrinsic value than high-ability Hispanic or White students, who had mean scores of 0.32 and 0.40, respectively. An explanation for lower science intrinsic interest may be that traditional science curricula are not personally relevant (Aikenhead, 1996; Barba, 1998; Bøe, Henriksen, Lyons, & Schreiner, 2011; Brickhouse, 1994). The idea that curricula should be relevant to all students is one of the key tenets of culturally responsive instruction (Barba, 1998; Ford, 2011). The intrinsic value scores showed large differences across groups; nonetheless, intrinsic value was not a significant predictor for any group at the $p = .05$ level. The lack of a connection between intrinsic interest in a subject and persistence is supported by

Holmegaard, Madsen, and Ulriksen (2012), who found that many Danish high school students claimed that a STEM subject was their favorite subject, yet avoided STEM majors in college.

Utility Value. Utility value measures how much students feel that science and mathematics courses are useful for future college or career plans. Such value is established when students are made aware of potential options for college or career and understand that mathematics and science coursework are important steps toward achieving such goals. The logistic regression models show that STEM utility value predicted the persistence plans of Hispanic students, but not for Black or White students. This difference indicates that for Hispanic students practical concerns of college and career take precedence over personal interest and enjoyment. Therefore, establishing the utility value of STEM is particularly important to motivating Hispanic students to take such courses. A lack of role models in the STEM fields is a barrier to the creation of utility value for STEM in Hispanic students (Hines, 2003). Interestingly, the values of STEM utility did not vary as much between race/ethnicity groups as some other variables did (Table 6), but STEM utility was predictive of persistence plans only for Hispanic students, and there was a large difference in mean STEM utility between Hispanic persisters and nonpersisters (Table 4).

Attainment Value. Higher science attainment values predicted the persistence plans of all students. This finding is supported by the previous studies of several researchers. In this study, science attainment value is based on how well the student's perception of the domain of science fits with the student's own identity. Aikenhead (1996) found that only 5–15% of students had a strong, positive sense of science attainment value and that this distinguished potential future scientists from other students. Oyserman and Destin (2010) explained differences in academic attainment as related to preferences for identity-congruent actions to identity-incongruent actions. Students who believe science and mathematics are identity congruent will have a higher attainment value for these courses and be more likely to plan to persist. Aschbacher, Li, and Roth (2009) also documented strong relationships between aspirations, persistence, and identity in their longitudinal study of a diverse sample of high school students. Furthermore, Bøe et al. (2011) related the problem of declining rates of STEM career choice to an increased focus on the occupation as an expression of identity and the fulfillment of the self that exists in more developed countries. Thus, the findings of the present study and previous research support the conclusion that students may not be willing to consider careers for which the characteristic traits are dissonant with desired personality traits that are part of their identities.

Some students may be more willing to consider careers that do not align well with their preferred identities because of sociocultural differences. The degree of willingness to deny the desires of the individual in favor of the needs of the group has sociocultural origins. Ford (2012) described cultural differences between Blacks, Hispanics, and Whites that included variations in views of the importance of a unique personal identity or of the importance of service to the community. These cultural views create differences in how much students are willing to compromise their preferred identity to conform to the expectations of a STEM identity. For example, students from less developed countries may be more willing to adopt STEM identities and pursue such careers.

Science attainment value measures the degree to which the student identified himself or herself as a science person and is identified by others as a science person.

In this study, significant differences in science attainment value were found favoring White students over Black or Hispanic students and favoring persisters over nonpersisters in all groups. Furthermore, the level of science identity was significantly related to persistence for all students. Therefore, methods to improve identity congruence of STEM and desirable student identities should be of interest to educators. Such methods should address changing school STEM curricula to increase the emphasis on qualities that are valued by students because these qualities are congruent with students' identities.

Cost. In this study, cost is the student's assessment of how much engagement in mathematics and science coursework will preclude other activities, require excessive effort, or affect relationships with peers. Cost was not found to be a significant predictor of persistence for any group. The White group had a slightly more positive sense of cost than the Black or Hispanic groups. However, this measure of cost references a more immediate cost—how time spent working on mathematics and sciences courses interferes with more desirable activities and may yield negative reactions from peers—compared to the more long-term social cost of adopting a stigmatized identity. Indeed, Holmegaard et al. (2012) found that students' avoided STEM identities that were in conflict with their ideal identities, and this was a reason why these students who claimed STEM subjects as their favorite subjects did not pursue STEM degrees. In effect, this represents a different type of cost, and a concern for entering an occupation that may not lead to self-fulfillment compared to a concern for the reactions of others. The operationalization of cost in the present study is aligned with the Eccles et al. (1983) model. However, it may be that long-term social costs are more relevant to occupational choice decisions than the immediate cost measured in this study.

IMPLICATIONS

The STV components of science attainment value, science intrinsic value, and STEM utility value are predictive of planned STEM persistence, but these variables may operate differently in groups of Black, Hispanic, and White students. The separate models described in the present study are a first step in examining between group differences. Further analyses are needed to establish the statistical significance of between group differences. These models can provide guidance for the development of interventions that could increase the numbers of students who plan to persist in STEM. For all students, identity congruence is likely to be a consideration in STEM persistence plans. One implication of this finding is the need to find ways to increase the congruence between STEM identities and students' identities for all students. The second implication is that students need to be made more aware of the utility of science and mathematics courses in relation to their future goals for career and college. The third implication is that STEM teachers and curricula need to inspire interest in these subjects. In this section, recommendations are made for practice and future research.

Schools should encourage the development of science identity in high-ability students by incorporating culturally responsive teaching principles into science courses and gifted programs (Barba, 1998; Ford, 2011; Hines, 2003). Research has shown that minority students' interest in science was positively affected by the integration of culture into science (Hines, 2003). Barba (1998) explained that science teaching must be more harmonious with culturally syntonic variables. For example, science classes that emphasize individual competition and where grading is on a curve do not fit well with the learning styles of

culturally different students who prefer to work more collaboratively and develop extended networks of support among their peers.

The manner in which courses are taught is important to the recruitment and retention of students in the STEM disciplines. Science courses need to shift from a traditional purpose of "weeding out" students who are believed to be not capable of science (Aikenhead, 1996) to a more progressive purpose of inspiring interest, scaffolding learning for all students, and scouting for talent. To serve this new purpose will require the use of research-based principles of teaching and learning with established effectiveness that are also culturally responsive. Students must learn about the nature of science and how science knowledge is created so that they can realize that their own ideas are valuable. Teaching strategies that emphasize active learning and collaboration such as problem-based learning or inquiry are culturally responsive because students can investigate issues that are relevant to them and participate in building scientific knowledge. Introductory courses must be interesting and engaging to inspire students to continue studies in that discipline.

Minority students may consider science foreign because they do not learn about any scientists or inventors from backgrounds similar to their own or encounter scientists in their communities (Hines, 2003; Taningco, Mathew, & Pachon, 2008). These students may internalize the idea that they cannot perform science or may feel that they must lose their racial identity to be assimilated into the culture of science. Culturally responsive teaching methods can increase student interest in science courses and facilitate students' crossings between their own culture and the culture of science. Science instructors should reduce language barriers to learning by connecting science language and students' native languages to develop students' skills in making "border crossings" between the different worlds they navigate in life (Aikenhead, 1996; Cooper, 2011). The adoption of culturally responsive teaching practices will facilitate increased identity congruence between student identities and a science identity and science attainment value will increase.

Science and mathematics teachers should strive to inspire interest in their subjects and to engage all students through culturally responsive teaching practices. Some school settings discourage interest and passion in gifted students (Fredricks, Alfeld, & Eccles, 2009). Contexts that encourage interest and passion are characterized by teachers who model enthusiasm, courses and assignments that present adequate challenge, and tasks that are meaningful, varied, and cognitively complex (Fredricks et al., 2009). These characteristics will encourage high-ability students to continue studies in that subject. To increase the utility value of mathematics and science, providing students with information and advice about career options and the corresponding educational requirements is critical. Students need accurate information about STEM careers, and this information should be part of science curricula and high school career counseling. Schools can better support students through the provision of counselors and teachers who have similar backgrounds as their students. Furthermore, greater care must be taken to look for potential STEM talent in students and to encourage high-ability students to persist in developing their talents in mathematics and science.

In this study, models for the persistence plans of three groups of ninth-grade, high-ability students were developed and compared. Differences in the predictive models between race groups revealed different relationships among the predictor variables. Understanding these differences between groups of students may help educators to become more culturally responsive. The finding that science attainment value is the strongest predictor of persistence plans for all groups is not surprising based on previous research. This study provides quantitative evidence based on the analysis of a large, nationally

representative sample that complements the findings of previous qualitative research on STEM persistence. The group of high-ability students is similar to Aikenhead's "potential scientists" (1996, p. 15), and this analysis reveals that even many in this select group do not identify strongly with STEM and do not plan to persist. This problem is common to many highly developed and modernized countries (Bøe et al., 2011). Aikenhead posited that the subcultures of the lifeworlds of students and the subculture of science must be understood so that teachers can facilitate the border crossings of students between these cultures. Almost 20 years later, the data from this study imply that the situation that he described has not changed much and little progress has been made toward this end. The field of science education continues to struggle with reform efforts that appear to be in conflict with recent government mandates, driven by accountability for results without regard to the processes used to obtain those results (see Southerland, Smith, Sowell, & Kittleson, 2007). Previous quantitative studies of STEM persistence have focused on the number and level of mathematics and science courses that students take in high school. The findings of this study, taken with previous work in this area, imply that merely pushing students to take rigorous courses will not increase STEM outcomes. As Holmegaard et al. (2012) found, students who like such courses may still not pursue STEM majors. What is needed is to increase the compatibility of a STEM identity with the identities of our students.

APPENDIX A: QUESTIONS USED TO CONSTRUCT EXPECTANCY VALUE SCALES

Scale	Question	Responses
	Questions asked separately for mathematics/science	
	What are the reasons you plan to take more mathematics/science courses during high school?	
STEM utility	–Because will help get into college	Yes/no
STEM utility	–Because it will be useful in college	Yes/no
Mathematics/science intrinsic	–Because he/she enjoys studying mathematics/science	Yes/no
Mathematics/science intrinsic	–Because he/she is good at mathematics/science	Yes/no
Mathematics/science attainment	You see yourself as a mathematics/science person	4-point on Likert
Mathematics/science attainment	Others see you as a mathematics/science person	4-point on Likert
Mathematics/science self-efficacy	You are confident that you can do an excellent job on tests in this course	4-point on Likert
Mathematics/science self-efficacy	You are certain that you can understand the most difficult material presented in the textbook used in this course	4-point on Likert
Mathematics/science self-efficacy	You are certain that you can master the skills being taught in this course	4-point on Likert
Mathematics/science self-efficacy	You are confident that you can do an excellent job on assignments in this course	4-point on Likert

(Continued)

APPENDIX A: CONTINUED

Scale	Question	Responses
	Questions not asked separately for mathematics and science	
	If you spend a lot of time and effort in your mathematics and science classes . . .	
Cost	You won't have enough time for hanging out with your friends	4-point on Likert
Cost	You won't have enough time for extracurricular activities	4-point on Likert
Cost	You won't be popular	4-point on Likert
Cost	People will make fun of you	4-point on Likert

APPENDIX B: SCALES AND RELIABILITIES

Scale	Created by	Number of Items	Alpha
Cost	Researcher	4	.75
Mathematics attainment value	NCES	2	.84
Mathematics intrinsic value	Researcher	2	.68
Mathematics self-efficacy`	NCES	4	.90
Science attainment value	NCES	2	.83
Science intrinsic value	Researcher	2	.73
Science self-efficacy	NCES	4	.88
STEM utility value	Researcher	4	.87

The authors would like to thank Dr. James H. Stronge for his feedback on an earlier version of this article.

REFERENCES

Aikenhead, G. (1996). Science education: Border crossing into the subculture of science. Studies in Science Education, 27, 1–52.

Archer, L., Dewitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). "Doing" science versus "being" a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity. Science Education, 94(4), 617–639.

Archer, L., Hollingworth, S., & Halsall, A. (2007). "University is not for me—I'm a Nike person": Urban, working-class young people's negotiations of "style," identity and educational engagement. Sociology, 41(2), 219–237.

Aschbacher, P. R., Li, E., & Roth, E. J. (2009). Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. Journal of Research in Science Teaching, 47(5). 564–582.

Atkinson, R. D., & Mayo, M. (2011). Refueling the U.S. innovation economy: Fresh approaches to science, technology, engineering and mathematics (STEM) education (pp. 1–178). Washington, DC: The Information Technology and Innovation Foundation.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. Child Development, 72, 187–206.

Barba, R. H. (1998). Science in the multicultural classroom: A guide to teaching and learning (2nd ed.). Boston: Allyn and Bacon.

Bøe, M. V., Henriksen, E. K., Lyons, T., & Schreiner, C. (2011). Participation in science and technology: Young people's achievement-related choices in late-modern societies. Studies in Science Education, 47(1), 37–71.

Bonous-Harnmarth, M. (2000). Pathways to success: Affirming opportunities for science, mathematics, and engineering majors. Journal of Negro Education, 69, 92–111.

Brickhouse, N. (1994). Bringing in the outsiders: Reshaping the sciences of the future. Journal of Curriculum Studies, 26(4), 401–416.

Cooper, C. R. (2011). Bridging multiple worlds: Cultures, identities, and pathways to college. New York: Oxford University Press.

Eccles, J. S. (2005). Studying gender and ethnic differences in participation in math, physical science, and information technology. New Directions for Child and Adolescent Development, (110), 7–14.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. Educational Psychologist, 44, 78–89.

Eccles [Parsons], J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). Expectations, values and academic behaviors. In J. T. Spence (Ed.), Perspectives on achievement and achievement motivation (pp. 75–146). San Francisco: W. H. Freeman.

Enders, C. K. (2010). Applied missing data analysis. New York: Guilford Press.

Erikson, E. H. (1968). Identity: Youth and crisis. New York: Norton.

Ford, D. Y. (2011). Multicultural gifted education (2nd ed). Waco, TX: Prufrock Press.

Fouad, N. A., & Smith, P. L. (1996). A test of a social cognitive model for middle school students: Math and science. Journal of Counseling Psychology, 43, 338–346.

Fredricks, J. A., Alfeld, C., & Eccles, J. S. (2009). Developing and fostering passion in academic and nonacademic domains. Gifted Child Quarterly, 54(1), 18–30.

Heilbronner, N. N. (2011). Stepping onto the STEM pathway: Factors affecting talented students' declaration of STEM majors in college. Journal for the Education of the Gifted, 34, 876–899.

Hines, S. M. (2003). Multicultural science education: Theory, practice, and promise. New York: Peter Lang.

Holmegaard, H., Madsen, L. M., & Ulriksen, L. (2012). To choose or not to choose science: Constructions of desirable identities among young people considering a STEM higher education programme. International Journal of Science Education, 1–30, doi:10.1080/09500693.2012.749362.

Hosmer, D. W., & Lemeshow, S. (2000). Applied logistic regression (2nd ed.). New York: Wiley.

Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., et al. (2011). High School Longitudinal Study of 2009 (HSLS:09). Base-year data file documentation (NCES 2011–328). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved October 10, 2011, from http://nces.ed.gov/surveys/hsls09/hsls09_data.asp.

Lee, O., & Luykx, A. (2006). Science education and student diversity: Synthesis and research agenda. New York: Cambridge University Press.

Lewis, B. F., & Connell, S. (2005). African American students' career considerations and reasons for enrolling in advanced science courses. Negro Educational Review, 56, 221–232.

Lohman, D. F. (2005). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. Gifted Child Quarterly, 49(2), 111–138.

Lynch, S. J. (2011). Equity and U.S. science education policy from the G.I. Bill to NCLB: From opportunity denied to mandated outcomes. In G. E. DeBoer (Ed.), The role of public policy in K-12 science education (pp. 305–354). Charlotte, NC: Information Age.

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. Science Education, 95, 877–907.

Mau, W. C. (2003). Factors that influence persistence in science and engineering career aspirations. Career Development Quarterly, 51, 234–243.

NAGC. (2011). Redefining giftedness for a new century: Shifting the paradigm. Washington, DC: Author. Retrieved November 15, 2011, from http://www.nagc.org/index2.aspx?id=6404.

National Academy of Sciences. (2007). Rising above the gathering storm: Energizing and employing America for a brighter economic future. Washington, DC: National Academy Press. Retrieved September 1, 2011, from http://www.nap.edu/catalog.php?record_id=11463.

National Center for Education Statistics. (2011). High School Longitudinal Study: 2009 EDAT extract codebook. Washington, DC: Author. Retrieved October 10, 2011, from http://nces.ed.gov/surveys/hsls09/hsls09_data.asp.

National Science Board. (2010). Preparing the next generation of STEM innovators: Identifying and developing our nation's human capital. Washington, DC: National Science Foundation. Retrieved August 15, 2010, from http://www.nsf.gov/nsb/publications/2010/nsb1033.pdf.

National Science Foundation. (2012). Women, minorities, and persons with disabilities in science and engineering [web page]. Retrieved November 10, 2012, from http://www.nsf.gov/statistics/wmpd/2013/race.cfm.

Navarro, R. L., Flores, L. Y., & Worthington, R. L. (2007). Mexican American middle school students' goal intentions in mathematics and science: A test of social cognitive career theory. Journal of Counseling Psychology, 54, 320–335.

Oyserman, D., & Destin, M. (2010). Identity-based motivation: Implications for intervention. Counseling Psychologist, 38, 1001–1043.

Seymour, E., & Hewitt, N. M. (1997). Talking about leaving: Why undergraduates leave the sciences. Boulder, CO: Westview Press.

Simpkins, S. D., & Davis-Kean, P. E. (2005). The intersection between self-concepts and values: Links between beliefs and choices in high school. New Directions for Child and Adolescent Development, 110, 31–47.

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. Developmental Psychology, 42, 70–83.

Southerland, S. A., Smith, L. K., Sowell, S. P., & Kittleson, J. M. (2007). Resisting unlearning: Understanding science education's response to the United States' national accountability movement. Review of Research in Education, 31(1), 45–77.

Syed, M., Azmitia, M., & Cooper, C. R. (2011). Identity and academic success among underrepresented ethnic minorities: An interdisciplinary review and integration. Journal of Social Issues, 67(3), 442–468.

Taconis, R., & Kessels, U. (2009). How choosing science depends on students' individual fit to "science culture." International Journal of Science Education, 31(8), 1115–1132.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. Science, 312, 1143–1144.

Taningco, M. T. V., Mathew, A. B., & Pachon, H. P. (2008). STEM professions: Opportunities and challenges for Latinos in science, technology, engineering, and mathematics. Los Angeles: The Tomas Rivera Policy Institute. Retrieved November 20, 2012, from files.eric.ed.gov/fulltext/ED502063.pdf.

Thompson, L. R., & Lewis, B. F. (2005). Shooting for the stars: A case study of the mathematics achievement and career attainment of an African American male high school student. High School Journal, April/May, 6–18.

Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. Journal of Education for Students Placed at Risk, 12, 243–270.

U. S. Department of Education. (2008). 2007 Elementary and secondary school survey. Washington, DC: Office for Civil Rights.

Wigfield, A., & Wagner, A. L. (2005). Competence, motivation, and identity development during adolescence. In A. J. Elliot & C. S. Dweck (Eds.), Handbook of competence and motivation (pp. 222–239). New York: The Guilford Press.

Zarrett, N. R., & Malanchuk, O. (2005). Who's computing? Gender and race differences in young adults' decisions to pursue an information technology career. New Directions for Child and Adolescent Development, 2005(110), 65–84.

Science
Education

Inviting Argument by Analogy: Analogical-Mapping-Based Comparison Activities as a Scaffold for Small-Group Argumentation

BRANDON R. EMIG,[1] SCOTT McDONALD,[2]
CARLA ZEMBAL-SAUL,[2] SUSAN G. STRAUSS[2]
[1]*The North Carolina State University, Raleigh, NC 27695, USA;* [2]*The Pennsylvania State University, University Park, PA 16802, USA*

Received 19 December 2011; accepted 6 November 2013
DOI 10.1002/sce.21096
Published online 6 January 2014 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: This study invited small groups to make several arguments by analogy about simple machines. Groups were first provided training on analogical (structure) mapping and were then invited to use analogical mapping as a scaffold to make arguments. In making these arguments, groups were asked to consider three simple machines: two machines that they had built, used, and made measurements with and one that they had not yet studied. Finally, groups were to produce an argument in favor of the machine that worked most like another machine. Seven of these approximately 50-minute analogical-mapping-based comparison activities were given to 55 preservice elementary teachers working in 15 small groups over 7 weeks. When used as a scaffold for argumentation in small groups, these activities were found to generate a need for discernment, which allowed for simple machines and their parts to be understood in and connected to the context. © 2014 Wiley Periodicals, Inc. *Sci Ed* **98:**243–268, 2014

INTRODUCTION

In this study, we invited small groups of students to consider simple machines as analogues and to make arguments by analogy to learn about them. All simple machines are analogous in that they trade distance for force. We consider that asking students to make

Correspondence to: Brandon Emig; e-mail: bremig@ncsu.edu

an argument by analogy is a form of scaffolding for student learning in this area. Analogies are commonly used in science and in everyday life as thinking and communication tools (Clement, 1981; Dunbar, 2001). Inviting explicit comparison of analogous scenarios or concepts has been shown to promote learning in individuals (Clement & Brown, 2008; Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1980; Kurtz, Miao, & Gentner, 2001). However, relatively little research has been done on how analogies can affect group communication and argumentation (Bellocchi & Ritchie, 2011; Gadgil & Nokes, 2009). The contribution of the present study will be to provide discourse-analysis-based qualitative findings describing what this process of linking analogy and argumentation looks like, as well as how analogy scaffolds small-group argumentation.

Various physics concepts can be demonstrated with simple machines, including work, efficiency, friction, ideal mechanical advantage, and actual mechanical advantage (Hewitt, 1999; Roth, 2001). Although the machines look and function differently, they are all in fact analogous. Each allows for the trading of distance for force, usually with the effort force to move something being reduced and effort distance being increased. Past research has asked students to talk and make predictions about how a given simple machine will function, use the machine and make related measurements, and design applications for the machine (Glasson, 1989; McKenna & Agogino, 1998; Roth, 1996; Tucknott & Yore, 1999).

The rich, tangible context provided by student-constructed simple machines actually operating in the world has been shown to be a good way to encourage physics-based discourse, shared meanings, and use of deictics (i.e., context-dependent words such as here, there, this, and that, etc.). With these, students in groups can communicate by pointing at, using and refining simple machines, related understandings, and related language (Roth, 2001). Talking about, representing, constructing, combining, and designing applications for the machines can be good ways for students to learn technological- and engineering-design processes as well as the physical concepts related to them. Furthermore, simple machines require only basic mathematics. This helps students make predictions about how they will work (the required forces, positioning, etc.), which they can then test (Roth, 2001).

LITERATURE REVIEW

Analogies, for the purposes of this research, exist when relationships between dissimilar objects correspond or can be mapped back and forth, even though the physical components themselves are different (Gentner, 1983). Figure 1 (reprinted with permission; Tohill & Holyoak, 2000) provides an example of the process of analogical mapping. Although the *features* best correspond between the two people, if one turns to the relationships of objects in the scenario, one sees a different correspondence, that of *function*. Since the role of the person in the top scenario is to restrain the dog, he would best correspond to the tree in the bottom scenario. Other correspondences include dog to dog (both have the same role), and cat (top) to person (bottom) (both are being chased by the dog). Contrasting and mapping the scenarios as analogues makes the relationships (e.g., restraining, chasing, nonparticipant, etc.) in the scenarios salient.

Contrasting cases have the power to make salient certain features that might otherwise go unnoticed (Bransford, Brown, & Cocking, 2000; Bransford, Franks, Vye, & Sherwood, 1989; Marton, 2006). For example, Bransford et al. (1989) found that in looking at a picture of a single house, people are unlikely to notice features such as the width of the chimney or of the door. When comparing pictures of six houses, however, these features become more apparent. The fact that there is a great deal of alignment between the houses makes the small differences salient. In short, comparison (or contrasting) of cases makes for easier noticing.

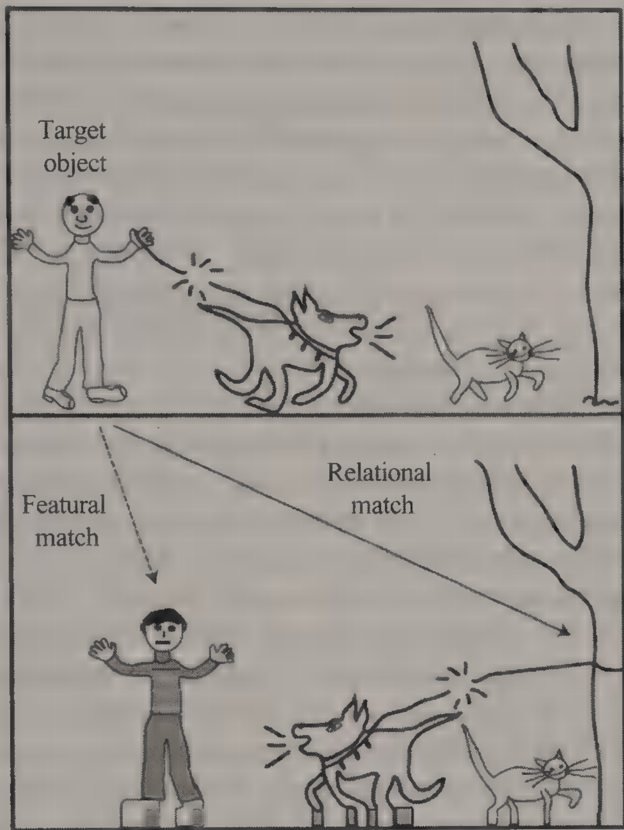


Figure 1. A training activity on analogical mapping.

Such comparisons can yield other benefits too. Mussweiler and Epstude (2009) found that study participants who engaged in comparison to an information-rich standard could make faster decisions or judgments about a new situation, required less information about a new situation, thought more about an information-rich standard, and were better able to carry out secondary tasks at the same time. The authors conclude that comparison carries with it efficiency benefits when dealing with a new situation or concept. Thus, the question "What is it like?" (i.e., what is like it that I may already know?) can be a faster and more effective way to learn about a new place, concept, or idea than "Describe it to me, please." These efficiency gains depend upon alignable features between the objects of comparison (Mussweiler & Epstude, 2009). But alignable features alone are not enough; rather, the *differences* make nuances more salient (Marton, 2006).

Analogies Naturally Scaffold Thinking

Analogical comparison can scaffold thinking to make a concept easier to understand or communicate. People use analogies spontaneously to think and communicate (Clement, 1988; Dunbar, 2001; Wong, 1993). For example, Clement (1981, 1988) invited experienced problem solvers to analyze a physics problem. Although they were not instructed to search their own background, Clement expected people to appeal to analogues from their past experiences to solve it. They did this, in fact, through an extended think-aloud process that produced correct reasoning via analogy. The author concludes that reasoning with analogies does not necessarily provide an immediate solution but has the potential to be effective over time. Furthermore, there is "reason to believe that some of these processes [of reasoning and conjecturing with analogies] are learnable, rather than being exclusively a product of genius" (Clement, 1981, p. 9).

Wong (1993) also found, among a class of preservice teachers, that "productive analogical reasoning can occur when the learners themselves assume primary responsibility for the task of generating, applying, and learning from analogies" (p. 1271). For instance, when participants were asked to explain the operation of a sealed syringe, they were found to appeal to self-generated analogies including the idea that air particles are analogous to BBs in a container, people moving in a room, or rubber balls inside the syringe (p. 1270). Participants reached into their past knowledge and experience to explain and understand the problem before them. This frequently resulted in the emergence of new questions, such as how to explain the fact that the syringe returns to its rest position when pulled out or even how to explain air pressure more generally.

In both Wong's and Clement's research, people were found to reason via an analogical reasoning process to understand and communicate ideas they did not fully understand. How have these findings been applied to science instruction? Many researchers have suggested that science students can be invited to use analogy as a *process* in the classroom (Duit, 1991; Else, Clement, & Rea-Ramirez, 2008; Wilbers & Duit, 2006). Analogies are commonly used as an explaining tool in science education (Brown & Clement, 1989; Coll, 2006; Fogwill, 2010; Heywood, 2002). For example, Brown and Clement (1989) argue for engaging "the student in a process of analogical reasoning in an interactive teaching environment" as opposed to "simply presenting the analogy in a text or lecture" (p. 237).

Various frameworks that have been developed to provide steps for the use of analogy in teaching and learning science (Else et al., 2008; Glynn, 1991; Treagust, Harrison, & Venville, 1998) all include language on analogy as an active process. For example, guidelines of Else et al. for working with analogies called for making "analogy as student-active as possible." The Focus, Action, and Reflection guide for working with analogies has an "action" phase that consists of "mapping of shared attributes" and "showing students where the analogy breaks down" (Treagust et al., 1998, p. 92). The Teaching With Analogies model also addresses the importance of process; its step four instructs students to "map similarities" (Glynn, 1991, p. 230). These frameworks all implicitly suggest the importance of comparing and contrasting, which has been shown to benefit learning.

ANALOGICAL COMPARISON PROMOTES LEARNING IN INDIVIDUALS

Comparison of analogous scenarios has been shown to be beneficial for learning in varied research. Kurtz and Lowenstein (2007) found that inviting participants to compare two problems significantly increased the performance on a new analogous problem (compared with a control that had participants read only one problem and its solution). Similarly, Podolefsky and Finkelstein (2007) found that students who used two analogical models for sound waves—an abstract one and a concrete one—were three times more likely to "reason productively" about sound waves than those receiving only one model.

Diehl and Reese (2010) found that learning improved when they invited learners to consider elaborated chemistry analogies. Clement and Brown (2008) also found evidence of learning through analogical comparisons. Specifically, they engaged a student in a process of analogical reasoning about a misconception that was overcome. They noted that "by establishing analogical connections between anchoring situations and more difficult ones, students may be able to extend the range of their valid intuitions to initially troublesome target situations" (p. 140).

There are, however, some potential problems in learning with analogies (Else, Clement, & Rea-Ramirez, 2003; Harrison & De Jong, 2005; Zook & DiVesta, 1991). For

example, individuals can make errors with analogies including overmapping (mapping correspondences where there are none), mismapping (mapping incorrect correspondences between elements), failure to map, and retention of base (better-known analogue) features (e.g., planets orbiting atoms) (Else et al., 2003, p. 8). Sometimes people have not benefited from analogies when they might have been expected to, without explicit guidance (Gick & Holyoak, 1983; Marton, 2006).

ANALOGY USE IN GROUP COMMUNICATION

Despite the benefits for individual learning, few studies have been done on the use of analogies for communication purposes in groups (Bellocchi & Ritchie, 2011; Gadgil & Nokes, 2009; Savinainen, Scott, & Viiri, 2005). Those that have focused on communication within group settings show encouraging results (Bellocchi & Ritchie, 2011; Fogwill, 2010; Mason, 1996; May, Hammer, & Roy, 2006; Oh, 2011; Savinainen et al., 2005). Bellochi and Ritchie (2011) evaluated how "analogy shapes classroom discourse" during analogy-writing activities (p. 771). Through video analysis of groups engaging with an analogy for electricity, researchers looked at how they made meanings as they navigated the space between the analogies (Bellocchi & Ritchie, 2011). The researchers found a particular kind of talk when one word or sign was made by students to apply to analogous scenarios. The researchers called this "merged discourse" (p. 786). They found that most instances of merged discourse were conceptually correct and that this type of discourse "was observed only during analogical activities" (p. 785). This work is important because when students merge their discourse, they are borrowing conceptual structures from both analogues. By allowing the analogy to be negotiated socially, the same word can be used, understood, and articulated from the perspectives of both analogues and by different individuals, thus benefiting learning.

Similarly, Oh (2011) had groups of students compare analogues. Having evaluated and transformed (e.g., graphed and otherwise organized) the data for four typhoons, students formulated explanations for another typhoon's path (which was different from the typical path) that not only drew upon the four analogues, but also went beyond, combining and extending elements from all. Although the four analogues were insufficient for explaining the anomalous path, they were useful in other ways. Oh suggests that more opportunities for this type of reasoning should be provided given that this is what "professional earth scientists are actually engaged in" (p. 429). Groups that are able to compare analogues have shown the ability to go beyond them in reasoning about new, related concepts.

These and other studies (Fogwill, 2010; Mason, 1996; May et al., 2006; Savinainen et al., 2005) describe the benefits of allowing groups the time and space to compare or even generate analogies. Fogwill (2010) found that "Imperfect students' analogies stimulated much more discussion than would any more perfectly mapped analogy provided in a text or a by a teacher" (p. 259). May et al., 2006found that even third graders in the United States could generate and modify analogies in response to others' arguments, and shortcomings perceived by others.

Small-group use of analogies has not been without its limitations, however. Yerrick, Doster, Nugent, Parke, and Crawley. (2003) found that although analogies provided a focal point for student group activity, without guidance students used them incorrectly or proposed others that were incorrect. This in part agrees with the work by Else et al. (2003), who found that individual students could overmap, mismap, or fail to map aspects of an analogy.

More research is necessary to evaluate the role of analogies in small-group communication and learning (Bellocchi & Ritchie, 2011; Gadgil & Nokes, 2009). This is important

given that analogy use and argument by analogy are both commonly used to aid in understanding and communication, even in everyday interactions (Brewer, 1996; Dunbar, 2001). Inviting students to make an argument by analogy holds promise in developing further understanding in this area.

The Importance of Argumentation and Scaffolding

To understand the nature and process of science, students must understand the process of argumentation that gives rise to scientific knowledge (Driver, Newton, & Osborne, 2000; National Research Council, 2007; National Research Council, 1996). Without understanding this process, students can perceive science as a body of facts that are self-evident and self-establishing. While engaging in argumentation, students should be doing activities that center on communicating, interpreting, and justifying scientific evidence with an eye toward understanding the scientific concept in question (Jimenez-Aleixandre, 2008). They can thus increase their understanding of the specific content to be argued as well as about science more generally. Although argumentation is important, it seems that learners do not discuss well or argue about what they do not understand (von Aufschnaiter, Erduran, Osborne, & Simon, 2008). They need to be supported in the argumentation process, and scaffolding is one way to do this.

The term scaffolding relates to the Vygotsky (1978) zone of proximal development, which is the theoretical space between what a novice can do without assistance versus what he or she can do with assistance from abler peers. Scaffolding of novices, or in this case science learners, means to temporarily support them to achieve a higher performance than they could achieve alone. It is expected that later, the learners will be able to perform unaided at the higher level (Mascolo, 2005; Wood, Bruner, & Ross, 1976).

Scaffolding need not be provided directly by another person who is present. Rather, it can be embedded into the environment to support a novice. One way of doing this is to structure and problematize science content for students (Reiser, 2004). To problematize content means to make it a situation in need of resolution as opposed to presenting it outright. (Note that this use of "problematizing" differs from other uses in such fields as technological design in which students come to form, understand, and articulate a problem to be solved.) Structuring content, on the other hand, involves making a task more doable by breaking it into steps. The two can be in tension at times; to structure too much can be to problematize too little, and vice versa. The goal of task-embedded scaffolding (i.e., structuring and problematizing) is to channel and focus student attention and action (Pea, 2004). When scaffolded well by problematizing and structuring content, student attention will be focused and channeled throughout the task.

With respect to scaffolding argumentation specifically, various strategies have been attempted, with encouraging results. One strategy is to scaffold students with computer software that can prompt students to attend to data in a specific order or that can arrange, represent, or graph data (Walker & Zeidler, 2007; Zembal-Saul, Munford, Crawford, Friedrichsen, & Land, 2003). Another is to provide students with various prompts or criteria for making their arguments (Jimenez-Aleixandre, 2008).

METHODS

Design Rationale

Scaffolding students' argumentation in science can improve their argumentation skills as well as their content learning (Walker & Zeidler, 2007; Zembal-Saul et al., 2003). Analogy

and argumentation, when put together, have the potential to yield benefits. Analogy serves as a scaffold for exploring new concepts in rich, mappable contexts. If groups are invited to make an argument by analogy, this work hypothesized, they may be scaffolded and better able to learn content through argumentation. Simple machines can be regarded as natural analogues that lend themselves well to this type of activity.

This project explores the scaffolding of argumentation with analogical-mapping-based comparison activities. Specifically, students in small groups were invited to make an argument by analogy about simple machines that are all, to various extents, analogous in function. This study specifically asked: What does it look like when argumentation and analogy are blended in the study of simple machines? How do comparison and analogical mapping affect communication and learning? What problems do students have? These questions were answered using the method of discourse analysis.

Context

This research took part in a science elective course for preservice teachers at a large research university. The three-hour-per-week course met twice weekly and had the dual aim of teaching science concepts via inquiry and designing inquiry-based activities. This research took place during the 8-week unit on simple machines in which simple machines were offered together as part of guided inquiry activities in a course for preservice elementary teachers to learn the science content, *and* to learn how to create and teach with guided inquiry lessons. Small groups of students built, used, modified, and made measurements of levers, pulleys, inclined planes, gears, and other simple machines. These activities asked small groups of about four students each to build the machines using Legos®. They were then used to lift something of known weight while students measured and recorded effort force, resistance force, and their corresponding distances. Various quantities could then be calculated. Questions embedded within the activity prompted students to summarize what they had learned about the simple machine under study. Finally, students designed a way to apply a combination of machines to lift a prescribed load. Class discussions, notebooking, and test questions generated reflections on the guided-inquiry, lesson-design principles employed.

Fifty-five students from three sections (14 groups of about four students each) elected to take part in the study, agreeing to be videotaped and have their written work used as data. The instructors were Betsy (two sections) and Mark (one section), both of whom had taught the course or a similar one for about one year prior. They helped to design the specific interventions.

Activity Development

When learning about simple machines, students are often given the chance to use the machines and asked to make measurements of and compare input and output forces and distances (i.e., effort force, resistance force [or load], effort distance, and resistance distance) (e.g., Hewitt, 1999; Roth, 2001). Then, in making, applying, and representing a given simple machine in various configurations (e.g., varied: loads, mechanical advantages) and/or varied simple machines (e.g., a pulley and a lever), students are guided toward a notion of mechanical advantage (as the multiplier through which effort and resistance forces and effort and resistance distances relate) (Hewitt, 1999; Roth, 2001).

The class in which this research took place did this as well. But from past semesters' experience, even with about 8 weeks to build, use, and compare simple machines, the

instructors had observed that students did not see the connections between all simple machines (i.e., all have calculable mechanical advantages, trade distance for force, have effort forces and distances, move a resistance through a distance, are to various degrees analogues to one another, etc.). Since the design of the course repeated not only simple-machine science content (and other science content) but also guided-inquiry lessons, the course allowed for machine–machine *and* lesson–lesson comparisons. Thus, it was unfortunate that the instructors felt that students were not achieving coherent understanding of the simple-machine content. For example, they were sometimes able to calculate ideal mechanical advantage for one machine but not for another, in spite of the similarities. The synergy hoped for from the repeated use of simple machines was not achieved. To address these concerns, this study focused on the simple-machine content of the course rather than the guided-inquiry pedagogical aspect. By inviting students to make an argument by analogy and to have us scaffold those arguments with an invitation (and training) to compare and map the simple machines as analogues, it was hoped that a more coherent understanding could be developed in which one machine's structure and function could support the students' learning of other machines.

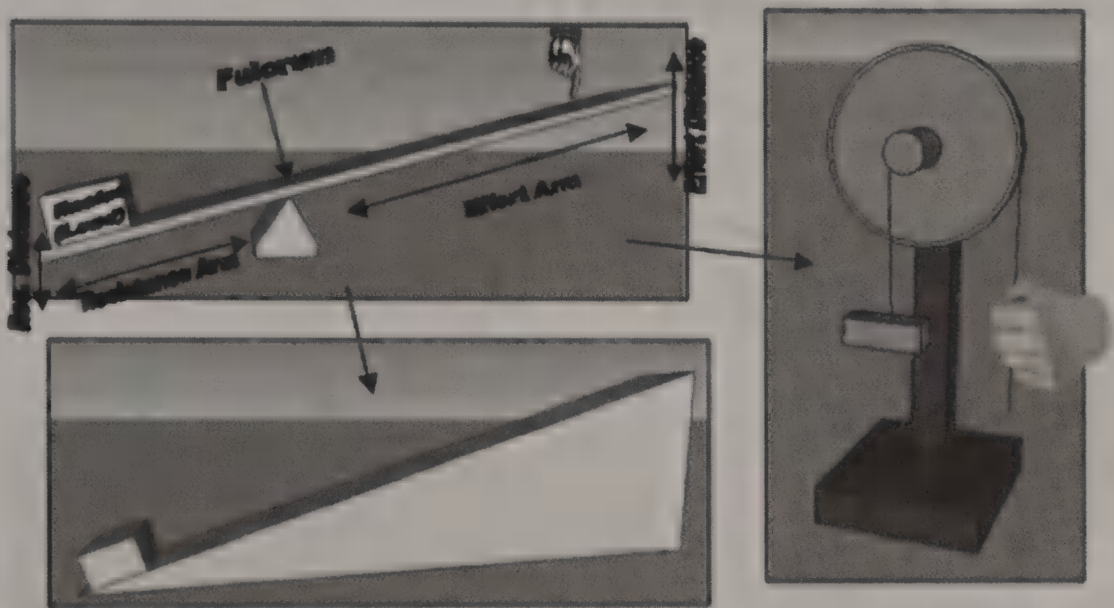
Activities were created to address this by inviting explicit comparison between simple machines by asking, "Is this (first) simple machine more analogous to this (second) simple machine or more analogous to that (third) simple machine? Why?" Figures 2–4 show actual handouts used for this research. During these activities, groups were asked to make an argument for a better analogue from among two possibilities: "Is science concept X more like possible analogue-concept A or possible analogue-concept B?" They had prior familiarity with two of the three concepts. The third was new to them. Before doing the activity in Figure 3, for example, students had already built, used, and made measurements with the pulley and the first-class lever, but they had not yet used or studied the wheel and axle.

The specific goal of this research was to learn and describe how small groups of students engage in argumentation when they are *scaffolded* with analogy-based comparison activities. By asking students to make an argument in favor of one simple machine as a best analogue, the simple-machine content was *problematized*, and by asking students to use analogical mapping to make that argument, it was *structured*.

The instructors and the principal investigator designed the activities to offer simple-machine juxtapositions that had the power to make important underlying concepts salient and invite student reflection on superficial similarities through the analogical mapping process and subsequent argumentation. For example, consider the shape- and position-based similarities between the lever and the inclined plane in Figure 4. They look similar, as do the wheel-and-axle and pulley in Figure 3. In both cases, however, these are not the best analogues. One must look deeper than shape or orientation. In spite of superficial shape similarities between the pulley and the lever (Figure 3), one looking to analogically map the axis of rotation of the wheel and axis would find a stronger analogue in the lever's fulcrum than anything the pulley might offer. While the pulley might seem to have a fulcrum (or perhaps more than one), these are quite different from those found in the other machines; they do not map well to the first-class lever, as they are not good analogues. Also, the wheel's radius affects the machine's mechanical advantage in a way that is analogous to the length of the effort arm of the lever—as both increase in length, the effort force required to lift is reduced. It was thought that analogical-mapping-based comparison activities such as this have the potential to scaffold this process of looking deeper in a systemic one-by-one way in which elements of one machine are considered in relation to another's.

A lever vs. a wheel & axle and an inclined plane

Which simple machine is most like the *lever* in the way it works? Why? *The wheel-and-axle* or *The inclined plane?*

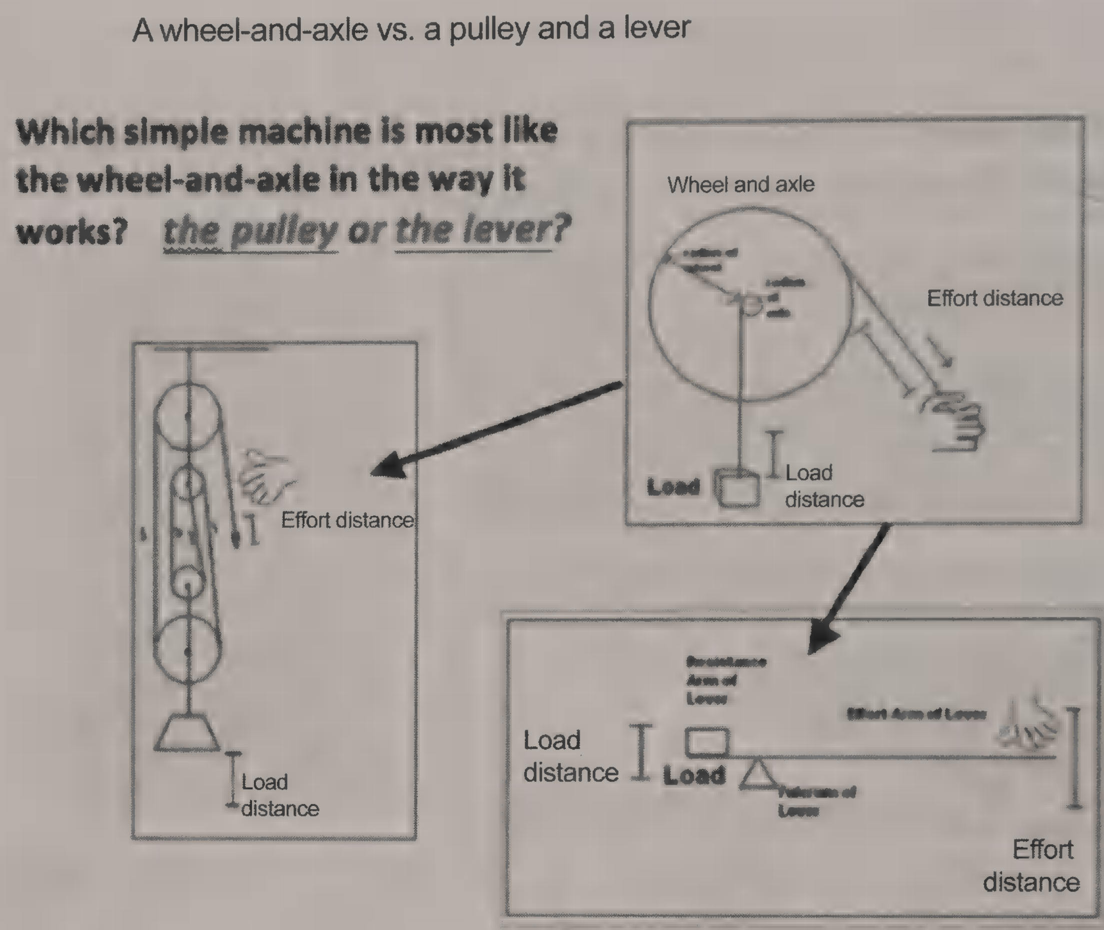


Lever	Inclined Plane	Lever	Wheel-and-Axle
Resistance Arm Length	**Resistance Distance**		

Figure 2. A lever mapped to a wheel and axle and an inclined plane.

Implementation

Student groups were trained to do analogical mapping by doing simple analogical-mapping activities (e.g., Figure 1) during the unit on simple machines. The training activities, each lasting about 15 minutes, were done in small groups, and then discussed as a whole class over two class periods. After the training, students in their groups completed, as in past years of the class, inquiry-based labs on simple machines in which they built, used, and made measurements. Interspersed with these, once per week, groups would do an analogical-mapping-based comparison activity, such as those in Figures 2–4, which would introduce a new machine not yet otherwise studied. (See the timeline in Table 1.) These

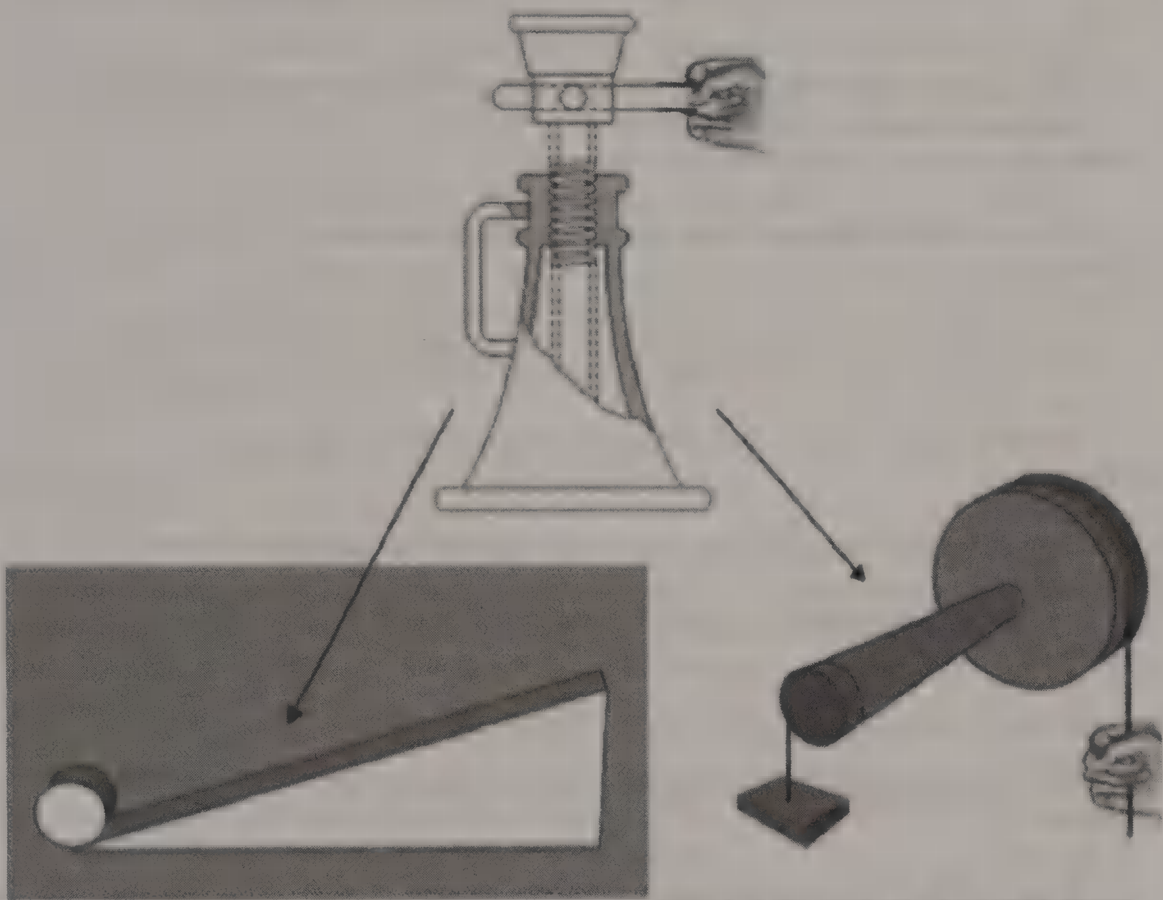


Wheel & Axle	Pulley	Wheel & Axle	Lever

Figure 3. A wheel and axle mapped to a pulley and a first-class lever.

took 30 to 60 minutes. After these activities, the small groups presented posters (roughly the same format as the handouts) with their analogical maps and final arguments to the rest of the class. Then, the whole class would discuss the groups' arguments and analogical maps and any problems they had doing them.

A screw-jack vs. a wheel-and-axle and an inclined plane



Jack	Inclined Plane	Jack	Wheel-and-Axle

Which simple machine does the screwjack work most like? The wheel&axle or the inclined plane? Use analogical mapping as learned in class.

Figure 4. A screw jack mapped to an inclined plane and a wheel and axle.

Data Collection and Analysis

Data were collected during the analogical-mapping-based comparison activities in the form of video and audio recordings, individual students' written analogical mapping tables, and posters of small groups. From among all 15 groups doing all activities, 48 hours, 38 minutes, and 6 seconds of video data were collected—an average of 43 minutes per group per activity. Because of equipment issues, one group's data were not transcribed for two of the activities. Only the video relevant to the research was transcribed. Side conversations

TABLE 1
Timeline for Interventions

I. Week 1
  a. Tuesday
    i. Participants Sought, Permission Forms Provided, Training 1—Comparative Argumentation Task 1—Dog Scenario
  b. Thursday
    i. Pretest
    ii. Students Build, Use, and Make Measurements with an Inclined Plane
II. Week 2
  a. Tuesday
    i. Training 2—Comparative Argumentation Task 2—General Science Concept
    ii. Comparative Argumentation Task 3—Inclined Plane vs. Screw
  b. Thursday
    i. Students Build, Use and Make Measurements with a 1st Class Lever
III. Week 3
  a. Tuesday
    i. Students Build, Use, and Make Measurements with All Classes of Levers
  b. Thursday
    i. Comparative Argumentation Task 4—Lever vs. Wheel and axle and inclined plane
IV. Week 4
  a. Tuesday
    i. Students Build, Use, and Make Measurements with Pulleys
  b. Thursday
    i. Comparative Argumentation Task 5—Pulley vs. Couch Lifters
V. Week 5
  a. Tuesday
    i. Comparative Argumentation Task 6—Wheel and axle vs. Pulley and Lever
  b. Thursday
    i. Students Build, Use and Make Measurements with Pulleys 2
VI. Week 6
  a. Tuesday
    i. Students Build, Use, and Make Measurements with Gears
  b. Thursday
    i. Test Review Discussion
VII. Week 7
  a. Tuesday
    i. Comparative Argumentation Task 7 Part 1 of Unit Test—Screw Jack vs. Inclined Plane and Wheel and Axle
  b. Thursday
    i. Part 2 of Unit Test on Simple Machines

longer than about five utterances and long periods of silence were not transcribed. Teacher comments and directions to the class were transcribed only once (as opposed to on each group's recording). This resulted in 24 hours and 21 minutes of transcripts. Transcripts were made by the principal researcher in StudioCode® and then pasted into Microsoft Excel.

The idea units within "reasoning sequences" (Pontecorvo and Girardet, 1993) were used as the unit of analysis for the transcripts. These are parts of the student argumentation in which "particular epistemic actions (or subactions) are performed" (p. 370) and only one thing is discussed. They last from two utterances (many seconds) to several dozen (several

minutes). Within these reasoning sequences, Pontecorvo and Girardet (1993) offer the term "idea units," which they refer to as "the smallest units in which the discourse is analyzed" (p. 370). An utterance may have zero to several idea units. Reasoning sequences were identified and highlighted in all transcribed data. These were reviewed by two instructors and the principal researcher together over several meetings to determine any patterns that could be labeled with codes. Next, they were shared with, critiqued, and modified by the other authors of this paper. Finally, they were applied to all data and evaluated again by the researchers. Disagreements were resolved through discussion.

The importance of discernment was noted by the researchers after reviewing videos and transcripts for early activities. Discernment for our purposes meant differentiating one thing (in this case one simple-machine element) from another. When it was not present, miscommunications between group members usually resulted. When it was present, communication was less problematic.

RESULTS

Our guiding research questions were as follows: What does it look like when argumentation and analogy are blended? How do comparison and analogical mapping affect communication and learning? What problems do students have? It was found that inviting groups to make arguments by analogy with simple machines called upon groups to do the following:

1. discern definitions and descriptions for simple-machine elements (parts, components, and related concepts such as effort force or resistance distance), which were important to be successful in making arguments by analogy;
2. go beyond superficial features of the machines in their argumentation to deep structural principles.

Most of the groups' overall machine-level arguments (machine X is most analogous to machine Y) were correct. (Note that some activities did not have only one correct argument (e.g., Figure 4).) Correct arguments ranged from 14 of 14 groups (activity in Figure 2) to 9 of 14 (activity in Figure 3). Since the sample size was only 15 groups, no significant effect sizes can be offered on student learning about simple machines. Also, pre- and posttests considered the 8-week unit as a whole (daily pre- and posttests were not done due to time limitations). In spite of the fact that most arguments were correct, no group found the analogical-mapping-based comparison activities to be without need for argumentation. The following discourse analysis focuses on this argumentation.

In early simple-machine-based activities, researchers noted that students were not sufficiently discerning with their words. For example, students used the word "effort" in lieu of a more discerning "effort force" or "effort distance." And they used "threads" (of a screw) instead of "thread length." In the first two activities based on the simple machine, none of the groups discern in any of the ways just mentioned. This resulted in miscommunications and misunderstandings. Analogical-mapping-based comparison activities require discernment to be completed well. To make an analogical correspondence between two simple machine elements, one needs to know what exactly those elements are, and why they correspond. It is easy to see how lack of discernment can lead to miscommunications and frustration in these activities. For example, if one student uses the word "effort" to mean "effort distance" and another uses it to mean "effort force," a communication problem will result.

Not only were misunderstandings caused by insufficient discernment noted in the transcripts, they were also evident in whole-class discussions after the activities. As a result, the

instructors and the principal investigator decided to discuss this with students as a recurring problem and invite students to become more discerning and pay attention to how they might do this more effectively in their small-group conversations. First, two reasoning-sequence excerpts of groups having a miscommunication due to insufficient discernment will be presented. Discourse analysis will show that these eventually ended in frustration either with an incorrect analogical correspondence or with a decision not to make one. Next, two reasoning-sequence excerpts from later activities will be given, of groups engaging in explicit discernment and making a correct analogical correspondence.

Miscommunications Resulting from Insufficient Discernment

The group in the following reasoning sequence had a miscommunication caused by insufficient discernment in an early activity in which they were comparing an inclined plane to a screw. In this case, this group considers a correspondence between the resistance distance of the inclined plane and an analogue element of the screw. By not discerning the differences between "resistance," "resistance force," and "resistance distance," communication about these features was hindered. Even physically pointing to these on the actual machines did not resolve the miscommunication issues, as the excerpt will show. First, however, some explanation is in order. See Figure 5. The upper and lower parts of Figure 5 are positioned similarly to show how they align analogically. (Note: the groups did not receive this labeled diagram.) The length of the screw shaft and the "resistance distance" as labeled on the inclined plane both represent the distance that a load would move up (imagine a block of wood moving up the screw while turning or conversely, the screw moving down into it). The force from friction and the need to split the wood would impart a resistance force on the screw. This means that although the load would travel along a larger distance (along the ramp length labeled "effort distance") at reduced force, it has really moved much less in terms of useful distance (just the vertical distance from the ground). This vertical distance or height of the inclined plane (or length of the shaft of the screw) can be called "resistance distance." The word "resistance" itself means essentially the same thing as load.

A screw vs. and inclined plane

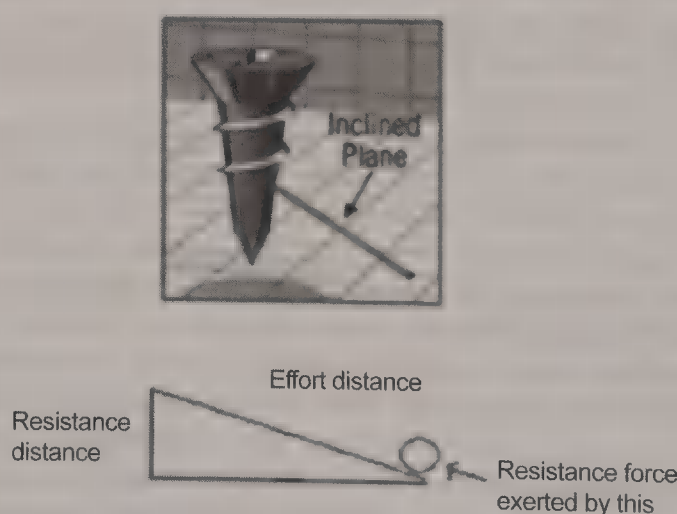


Figure 5. A screw shown with an inclined plane.

It is important to point out that the term "resistance distance," however, is different from the word "resistance." In the case of the inclined plane, the "resistance" would be the weight of the load at the bottom left. For the screw, the "resistance" would be something into which the screw was being turned—a wall for example. Both these resistances would exert a force on their respective machines, which can be called a "resistance force." Thus, it is important to distinguish between "resistance," "resistance distance," and "resistance force," as these are different things.

In the reasoning sequence below, group members Serena and Sheri use the word "resistance" to mean "resistance force," whereas Evan uses it to mean the "resistance distance." In the transcript, ideas relating to resistance are italicized for emphasis. The group begins talking about the screw:

00:20:16:47	Serena	I think the *resistance* is gonna be whatever it's going into.
00:20:22:08	Evan	The shaft would be—
00:20:22:87	Serena	But that's the *resistance*.
00:20:25:17	Evan	Is the shaft not the *resistance*?
00:20:31:11	Serena	I don't think so.
00:20:32:21	Sheri	I think this (points along *resistance distance* on inclined plane) would be like the shaft
00:20:36:38	Serena	Yeah
00:20:35:82	Evan	Why would that be like the shaft?
00:20:36:68	Sheri	I don't know. Cause it's constant.
00:20:38:59	Evan	Yeah but—the only reason thee—the threads are—have to do with the effort distance is cause they're going up the shaft. So, I feel like—the shaft would have to do with this piece (points to *resistance* distance on inclined plane)
00:20:50:77	Serena	But that's not gonna be—but that's the *resistance*
00:20:51:23	Sheri	But that wouldn't be *resistance*
00:20:52:40	Evan	That's not what the piece is called.
00:20:56:62	Serena	Yeah huhh [like uh huh]
00:20:54:98	Sheri	Yeah. It is.
00:20:55:49	Serena	It's *resistance*
00:20:56:94	Evan	That would also be the explanation.
00:20:58:06	Serena	But it wouldn't be *resistance*—the shaft isn't
00:20:59:15	Sheri	The wall would be resisting the threads.
00:21:03:45	Evan	How is it—
00:21:11:29	Evan	Why are you talking about wall?
00:21:12:76	Sheri	I don't know. Cause I don't see how anything else makes sense.
00:21:14:24	Serena	Like whatever it's going into it's gonna be resisting the effort.
00:21:17:94	Sheri	Yeah. And, I'm assuming it's going into a wall.
00:21:20:93	Evan	But that's not what it's at. There is no wall in this picture.
00:22:05:39	Sheri	(overlapping) I don't care what we write. I don't understand this.

Serena starts off stating the "resistance is gonna be whatever [the screw is] going into." This is partially correct. A wall or a board, etc., puts a *resistance force* on the screw. She and other group members, however, do not discern that "resistance," what they are saying, is different than "resistance distance" or "resistance force." Evan follows up with "The shaft would be . . . ". Although grammatically incomplete, he does bring up the "shaft." This is appropriate, since the shaft length of the screw corresponds analogically to the *resistance distance* of the inclined plane.

In this analysis, inferences must be made about some of the meanings intended. However, given the benefits of hindsight, correct scientific understanding, the fact that students seemed convinced of their respective positions, *and* that student argumentation positions are correct if interpreted in this way, these assumptions are reasonable.

Evan is influential in this argumentation, yet he is unable to hold sway without proper discernment between "resistance" and "resistance distance." Serena reiterates, "But that's the resistance." Evan, likely knowing that the *shaft length* relates to the *resistance distance* but insufficiently discerning, asks, "Is the shaft not the resistance?" Serena answers, "I don't think so." Note that he also is insufficiently discerning with the word "shaft." It is likely that he means "shaft length," as these would correctly correspond analogically.

Sheri continues, stating, "I think this (she points back and forth along the *resistance distance* on inclined plane) would be like the shaft." Unfortunately for Sheri, none of the other group members directed their attention to her physical pointing during her use of deictic language (context-dependent language including pointing; e.g., this, here, that, there). Serena agrees, stating, "Yeah." Again, this is nearly correct, but it again lacks discernment. Note that Sheri's use of "this" and her pointing along the inclined plane where the resistance distance is, suggest that she is talking about—but not saying—the *resistance distance*. Sheri's word, "resistance," is not sufficiently discerning. Had she said that "resistance distance" was like the "shaft length," this would have been correct and discerning. It is likely, since Sheri referred to "the shaft" of the screw and pointed to the *resistance distance* of the inclined plane, that she thought she was conveying this correct idea.

During the next few utterances, the opportunity for discernment was provided; however, it did not happen. Evan asks, "Why would that [resistance distance, as Sheri had mentioned] be the shaft?" Sheri states, "I don't know cause it's constant." This is an unclear response to Evan's question. Evan states, "Yeah but—the only reason thee—the threads are—have to do with the effort distance is cause they're going up the shaft. So, I feel like—the shaft would have to do with this piece [points to resistance distance on inclined plane]." This utterance has two important functions. First, Evan situates the new potential correspondence within another previously agreed upon one (not shown in this reasoning-sequence transcript)—that of the length of the screw thread and the length of the inclined plane (or effort distance). This attempt to give support to the new correspondence might have been effective. Evan even hedges somewhat with his use of a less-than-specific "the shaft *would have to do with* this piece." The "would have to do with" suggests the need for further discernment through argumentation. Evan still has not uttered the term *resistance distance*, though he has pointed along it on the inclined plane, but here again, as with Sheri's earlier use of deictic language, none of the group members direct their eyes to where Evan is pointing. Thus, although communication might have benefited because of it, it does not.

Evan believes the group is talking about *resistance distance*, or the distance a resistance is moved. Sheri and Serena believe the group is talking about *resistance force*, the force applied by the resistance, such as a wall. At utterance 00:21:11:29 Evan questions their right to assert that a wall exists when it is not pictured. The argumentation continues on this idea for about 54 seconds (not shown due to space limitations). And finally, the reasoning sequence ends in dissatisfaction when Sheri states, "I don't care what we write. I don't understand this."

The frustrating ending for the group is unfortunate, especially since there were many assertions that would have been correct had more discernment been used. Nonetheless, given the near correctness of their assertions, it is all but certain that all three members maintained key correct understandings (a fourth member was present and paying attention but did not participate during this reasoning sequence).

This reasoning sequence is representative of a class of such miscommunications caused by insufficient discernment that occurred in every one of the 15 groups. Some were longer. Some were shorter. But the key elements were the same: lack of discernment in communication causes a miscommunication. While doing the same activity, all but four groups had nondiscernment-caused miscommunications nearly the same as the one analyzed here (i.e., dealing with resistance). All groups, however, had at least one miscommunication due to lack of discernment.

The use of "threads" to mean thread length was common. All groups used the nondiscerning term "threads" (as opposed to thread length); two groups did manage to use "thread length" after being guided to the term by an instructor after posing questions with the activity due to difficulty. Still, this did hinder communication. For example, a person in another group wondered whether "larger" threads would make a screw easier or harder to turn. This miscommunication began with the following utterance:

> Madeleine Oh no—if like the threads were larger, it'd take more effort to screw it into something.

Madeleine took "larger" to mean "a thicker screw," whereas another group member took it to mean longer threads (and thus finer) on the same screw shaft, which would reduce the effort force; and another member took this to mean farther apart (requiring more effort force). The word "larger" is insufficiently discerning. Only four groups (of 15) made this particular correspondence between "threads" and "effort distance" without problem or apparent miscommunication. Nonetheless, these four groups show that some groups *did* accept an analogical correspondence even with an insufficiently discerning word choice without much apparent difficulty.

These activities and their use of analogical mapping and simple machines easily allowed for and permitted the determination of nondiscernment-based miscommunications. Many aspects of simple machines had similar or closely related terms (e.g., resistance distance and resistance force, effort arm, effort distance, and effort force) that, when combined with nondiscerning word choice, can allow space for miscommunication. This *limited* space combined with the benefit of hindsight, complete transcripts, and knowledge of correct analogical correspondences, made identification of such miscommunications possible as shown in the previous analysis.

Instruction Toward Discernment: Prompting a Shift

During the activities, the instructors and the principal investigator were walking around the room. From listening to student small-group argumentation and the subsequent whole-class discussions, it became evident that lack of discernment was a problem. In the first activity on simple machines, as mentioned earlier, all groups made a correspondence between the "threads" (instead of "thread length") of a screw and the "effort length" on their posters. Not even the two groups that were helped by instructors used "thread length" on their posters. Instructors addressed this. For example, after students shared their posters with their analogical maps and arguments with the whole class, the instructors commented that all groups had used the word "threads" when making a correspondence to the "effort distance" (or in some cases just "effort"). They stated on that day that "thread length" would have been more appropriate as "threads" are a concept or idea, whereas "thread length" can be measured. Instructors made similar mentions to individual groups about the need

to differentiate between "effort distance" and "effort force" and other related terms during following activities as well.

Reasoning Sequences from Activities Showing Explicit Discernment

In the following section, two reasoning-sequence excerpts of groups showing explicit discernment will be shown and analyzed. In the first reasoning sequence, the group argues about a correspondence between the effort forces in both the lever and the wheel and axle. Although groups had measured and used the concepts they are about to discuss (effort force and effort distance) in earlier lab activities, the concepts remained problematic. The activities, as will be shown, invite the group to discern the difference. The handout for this activity is shown in Figure 2. Interestingly, the term "effort forces" relates to all simple machines in the same way (i.e., the distance over which the hand applies the input force). Nonetheless, the group needed to engage in argumentation to confidently make the connection and locate these on the different machines. The group was able to do this due to the explicit discernment made between "effort force" and "effort distance."

Beth begins the reasoning sequence by suggesting a relationship between the "hand pulling down" and the "hand pushing down." (Figure 6 provides a superimposed lever and wheel illustrating the concepts in this reasoning sequence.) She states, "Well.—Let's stick to the lever and the wheel-and-axle—because I think the hand pulling down could be equal to the hand pushing down so maybe the effort. See now what do you call that?" In this last phrase, Beth explicitly directs the group's focus to what they should call "that." Here again we see the use of deictics in Beth's "that." The rich context provided by the simple machines comes to the fore. At the beginning of the reasoning sequence, the contextual "that" becomes a seed for further discernment when combined with her question on what to call it. In her utterance, she had also used the word "hand" to center her thoughts on "effort." Consider the rest of the excerpt given below.

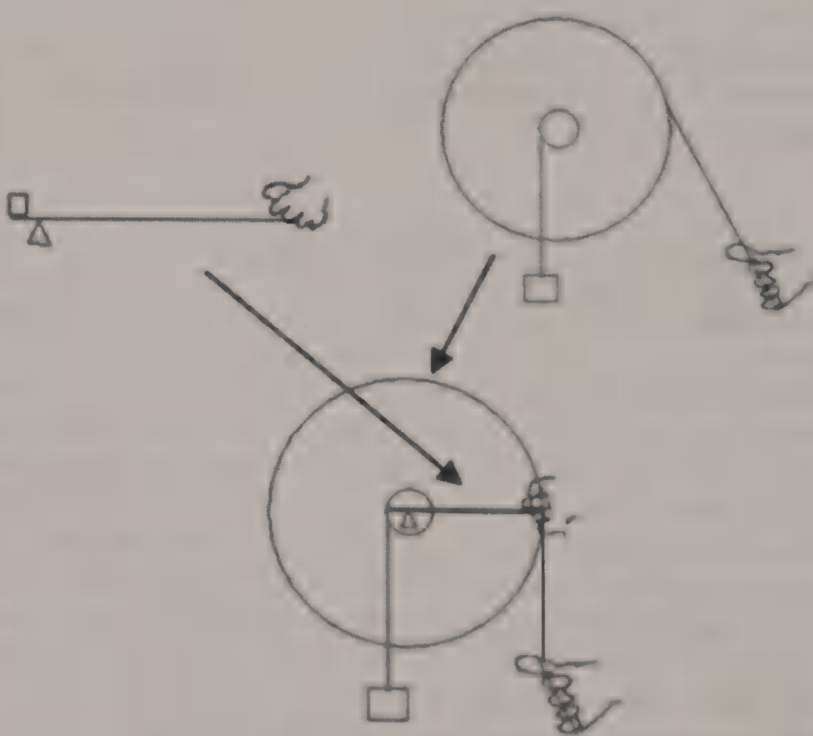


Figure 6. A lever superimposed on a wheel and axle.

00:12:15:35	Beth	Well—Let's stick to the lever and the wheel-and-axle—because I think the hand pulling down could be equal to the hand pushing down so maybe the effort. See now what do you call that?
00:12:32:42	Bree	Effort force.—Right?
00:12:32:58	Melissa	Yeah.
00:12:34:02	Beth	Wow. Look at you!
00:12:36:52	Bree	Yeah. I know. Sometimes I get them right.
00:12:38:00	Beth	What would you call that on the other one, effort force?
00:12:42:81	Bree	Yeah. Actually.
00:12:46:56	Beth	I'm just gonna put hand coming down, hand pulling in parentheses.
00:12:52:67	Bree	They're both effort forces right?
....		
00:13:28:25	Dory	But is there a difference between effort and effort distance? I think there is.
00:13:33:72	Melissa	There's effort force and effort distance
00:13:32:97	Bree	Yeah. So I don't know which one it is.
00:13:34:11	Melissa	Well that's effort distance (points along effort distance). Effort force is what—the force that it takes to pull the thing up the effort distance.

Picking up on Beth's question, Bree offers "Effort force. (2 s) Right?" The next three utterances convey confidence in Bree's assertion. Melissa follows with, "Yeah." Next, Beth adds, "Wow. Look at you!" Bree then states, "Yeah. I know. Sometimes I get them right." The "effort force" has been correctly discerned and the term appropriated and the group members know it.

The next few utterances serve to verify the fact that the label applies to both hands, not just one. (The "hands" can be seen in Figure 2.) Beth asks, "What would you call the other one, effort force?" And Bree says, "Yeah. Actually." Then Beth apparently hedges somewhat; perhaps the words "effort force" might be insufficiently discerning. She says, "I'm just gonna put hand coming down, hand pulling in parentheses." Bree, in spite of her previous utterance, solicits further verification; "They're both effort forces, right?" Combined with the previous several utterances, perhaps this utterance was made to further convince other group members of the correctness of her idea, or perhaps it was made to refute Beth's need for the additional detail "in parentheses."

After about 36 seconds of unrelated dialogue, Dory, who had not yet spoken in this reasoning sequence, sought additional discernment in asking, "But is there a difference between effort and effort distance? I think there is." Melissa responds, "There's effort force and effort distance." Interestingly, once the discernment between "effort distance" and "effort force" had been offered, Bree seems to question her prior utterances stating, "Yeah. So I don't know which one it is." It is possible that when she had stated "effort force" in her earlier utterance that she did not realize there was an "effort distance". Or, maybe she simply did not mentally juxtapose the two. More likely, however, she did not know what exactly was best represented by the hand (see Figure 2). Melissa reaffirms her initial assertion stating, "Well that's effort distance (points along effort distance). Effort force is what—the force that it takes to pull the thing up the effort distance." This utterance comes full circle in answering the question posed by Beth in the first utterance: what to call it. Melissa offers an important final discernment with deictics (i.e., pointing and using "that's") between "effort distance," "effort force," and just "effort," the earlier used word, which is unclear.

The comparison activity and related instruction scaffolded the students' discourse toward the discernment of a definition for effort force in context as can be seen in this reasoning sequence. First, Beth began the sequence using context-specific language and the word "hand" in an attempted correspondence that suggested that these might relate to "effort." She then asked for more discernment around the word "effort." Bree then offered "effort force." Dory explicitly next asked the group to discern between "effort" and "effort distance." Melissa offered further discernment to Dory's "effort," stating there's "effort force" and "effort distance." Finally, Melissa offers a clearly discerned and pointed out definition of both "effort force" and "effort distance."

The analogical-mapping-based-comparison activities created a need for discernment and a context within which it could take place. Both simple machines offered a perspective from which to view "effort force" (and "effort distance"). In addition, the coconstructions made possible between group members allow for easy changes back-and-forth between those two perspectives.

A Second Example of Discernment

The next excerpt, from Haley, Nathan, Audrey, and Jenn, shows another example of explicit discernment. The group is engaged in argumentation on the same activity (Figure 2) as the group in the previous excerpt. They are attempting to find simple-machine element correspondences between the lever and the wheel and axle.

Audrey begins by asking, "The fulcrum and—the—the thing–isn't that the same as the (points to wheel-and-axle) pivot point. Not pivot point. The center of the thinger." Consider the transcript below.

00:10:37:00	Audrey	The fulcrum and—the—the thing—isn't that the same as the (points to the wheel-and-axle) pivot point. Not pivot point. The center of the thinger
00:10:46:59	Haley	Yeah
00:10:47:81	Audrey	Wheel and axle
00:10:48:63	Haley	So the fulcrum—Should we just call it that the thinger? (laughs)
00:10:57:11	Audrey	There has to be a smarter word for that. Center thingy. Come on Nate. We need your big words here
....		
00:35:22:13	Haley	Ok. The fulcrum and the center are the same because—
00:35:26:44	Audrey	Because that's like the pivot point of the—machine

Audrey's utterance serves two functions. First, she introduces a possible correspondence between the lever's fulcrum and the wheel. Next, she questions her own use of "pivot point" as an acceptable term to make a correspondence to the lever's fulcrum. The word "thinger" combined with deictic pointing also promoted the need for discernment early in the reasoning sequence. This questioning makes it acceptable to the rest of the group to engage in discernment around finding a better term. Her questioning of the term also allows her to save face should a better term emerge from further argumentation. Haley agrees, stating, "Yeah." Audrey tries to clarify with, "Wheel and axle." It is not apparent whether this was a question or a statement.

With Haley's "So the fulcrum—Should we just call it that the thinger? (laughs)," the dialogue next turns explicitly toward discernment. Clearly, "the thinger" is insufficiently

specific to correspond with the fulcrum. Audrey states, "There has to be a smarter word for that. Center thingy. Come on Nate. We need your big words here." The contiguous reasoning sequence ends here. The group did, however, take up the matter in a follow-up reasoning sequence when attempting to write final choices on poster paper for sharing approximately 25 minutes later, offering a final two utterances. Haley states, "OK. The fulcrum and the center are the same because . . . " to which Audrey responds, "because that's like the pivot point of the . . . machine." The simple machine element in question would best be called the axis of rotation. The members of the group likely had heard this term at some point before. Regardless, their definition was exact and well discerned, pointed out, and referred clearly to the axis of rotation in spite of the use of different words.

Although the argumentation in the end yielded a product much like the one in the first utterance, the "center" of the wheel and axle or the "pivot point" of the wheel and axle are specific and unique enough to not be confused with any other element. Therefore, it is considered that discernment took place between the words "thinger," "pivot point," and "center." And although "pivot point" ultimately was adopted, the other terms, as well as the physical pointing, added to the communication and discernment process.

DISCUSSION

Inviting groups to make arguments by analogy with simple machines called upon groups to do the following:

1. discern definitions and descriptions for simple machine elements (parts, components, and related concepts such as effort force or resistance distance), which were important to be successful in making arguments by analogy.
2. go beyond superficial features of the machines in their argumentation to deep structural principles.

Reasoning sequences showed that the invitation to use analogy and analogical mapping scaffolded groups in their argumentation toward discernment of definitions and descriptions. Encouraging analogical argument, as has been done here, has been found to promote discernment. Discernment has played a key role in doing analogical-mapping-based comparison activities. Furthermore, students' discourse went beyond superficial appearances toward deeper structural and functional principles. Concepts as opposed to appearances were discussed. This is emphasized more by what was not said than what was said.

In the first example, insufficiently discerning word choice for the description of a simple-machine element that led to miscommunication and a frustration marked an ending to the reasoning sequence. In the final two examples, explicit discernment was noted early in the reasoning sequences, when after nondiscerning terms had been uttered, the need for discernment (e.g., "what would you call that?," "how would you say that?," etc.) or discerning terms that build on what was already said were offered (e.g., "I'd call that a . . . ," "In this case it would be effort force, not just effort," etc.). The reasoning sequences with discernment ended in a confident choice for an analogical correspondence.

The utterance, "[e]ffort force is what—the force that it takes to pull the thing up the effort distance," from reasoning sequence three, came at the end of the sequence and shows that students did gain understanding. This is all the more compelling since students had measured dozens of examples of effort distances and effort forces (with a ruler and a spring scale), recorded them, and used them to answer questions. An argumentation *process* in the group still needed to take place to arrive at this understanding. If students had not

earlier reached an understanding on what "effort force" and "effort distance" were, then what about the elements of simple machines that had not earlier been measured? What about the fulcrum of a lever or the axis of rotation of a wheel-and-axle, for example? It is safe to say, for reasons similar to those just discussed, that these were not well understood prior to the analogical mapping activity. They needed to be discerned and defined, and this was not unproblematic. In the final reasoning sequence, Haley and Audrey offer the following coconstruction at the end, "OK. The fulcrum and the center are the same because—because that's like the pivot point of the—machine." The collaborative process of analogical-mapping-based argumentation had a product—understanding—articulated in that utterance. Students considered elements of simple machines, labeled them, and argued about them to make an analogical correspondence.

Through the argumentation-by-analogy process, multiple machines and individuals provide multiple perspectives, which can benefit learning by offering different understandings and points of view of a given concept. Without problematizing and structuring student learning in this way, a fulcrum might otherwise be just a vocabulary word to be memorized. Through discernment, by contrast, a fulcrum buttresses and contextualizes an axis of rotation, and an effort arm buttresses and contextualizes a wheel's radius.

The representative reasoning sequences analyzed have shown that the analogical-mapping-based activities both *problematized and structured* the simple-machine content for students, the very components of scaffolding adopted for the purposes of this research (Reiser, 2004). The excerpts included showed problematization, since the analogical-mapping-based comparison activities and their requisite final arguments (e.g., machine X works more like machine Y) generated a need for the small groups to identify, and label in proper order, the elements of simple machines (e.g., fulcrum, effort distance, effort arm, resistance force, etc.) to make analogical correspondences between those elements. The analogical-mapping-based comparison activities *structured* content as well. Students were provided a systemic approach to identify and label elements of simple machines. Thus, students were able to use and point to one simple machine as a reference point to investigate another. On the basis of the discourse analysis here, it is presumably easier to state that a lever's "fulcrum" is like a "pivot point" of a wheel-and-axle and point these out (as the group used in the last reasoning sequence analyzed) than it would be to come up with an absolute definition for a fulcrum. By making the task doable in this way, the activity is structured for students.

The instructors, of course, played an important role in the activities. First, they offered analogical-mapping training to students. This was necessary in order for students to first have a good idea of how to identify elements of a scenario and make and explain an analogical correspondence. Also, the activities did not sufficiently structure the content by themselves, as indicated by the lack of discernment in some activities. Therefore, instructors also socially scaffolded students by making them aware of the need to be more discerning. Given the natural need for discernment to accomplish the activities, they provide good environments for learning to be discerning.

What was not discussed was also important. For example, no groups were found to attend to extraneous features such as color or relative size. Also, despite the same superficial positioning of the inclined plane and the lever shown in the handout in Figure 2, no groups made this correspondence. Superficial appearances were simply not important in making correspondences. All of the reasoning sequences analyzed in this paper showed students dealing with concepts that were pertinent to the task of analogical mapping and argumentation. This is further evidence that student attention was focused and channeled over the time on task, which is our definition of scaffolding.

CONCLUSION AND IMPLICATIONS

By blending argumentation and analogy, this research has provided a way to make analogy a process, as many researchers have recommended for science education (Brown & Clement, 1989; Else et al., 2003; Gentner et al., 2003; Glynn, 1991; Heywood, 2002; Nashon, 2004; Theile & Treagust, 1991; Treagust et al., 1998; Wilbers & Duit, 2006). Inviting argument by analogy offers a way to scaffold argumentation for students in order for them to learn the content. The simple machines used in this research, although not literally the same, had analogous structures. Contrasting and comparing them through argumentation made features important, noticeable, and salient. Alignable structure is not enough, however; the differences needed to be explored. Analogical mapping allowed for this.

Various frameworks for using analogy in instruction in science education are in alignment in the way they stress the *process of analogy* (Else et al., 2003; Glynn, 1991; Nashon, 2004; Treagust et al., 1998). This study has contributed to the knowledge of what can happen during that process, what some of the problems are (i.e., nondiscernment), and possible ways to improve the use of analogy (e.g., encourage discernment, attend to differences between analogues, inviting arguments based on analogy, and involved groups of students).

Although research has shown that reasoning and communicating by analogy is common and effective (Brewer, 1996; Clement, 1981; Dunbar, 2001; Wong, 1993), few studies have focused on how analogies can influence classroom discourse (Bellocchi and Ritchie, 2011; Gadgil & Nokes, 2009). This study showed the argumentation process with its deictic and coconstructed utterances playing out in the rich simple-machines-as-analogues context, requiring discernment and shared meanings.

This study also contributes to the literature on the argumentation process in science education. (Before discussing this further, however, it is important to note that given the conceptual nature of the arguments, categories such as evidence, claims, warrants, backing, etc., based on the work of Toulmin (1958) [still commonly used] are not present.) For example, Jimenez-Aleixandre (2008) suggests that students should be "active producers of justified knowledge" and to accomplish this, student roles should include generating products or answers, choosing among competing explanations, backing claims, distinguishing good from poor arguments, talking science, and persuading peers (pp. 96-7). She also suggests that students should "generate products or answers." In the present work, students did this in two ways. First, they made analogical comparisons between elements of simple machines (e.g., a fulcrum of a lever is like the axis of rotation of the wheel-and-axle). Second, they made a principal argument between two simple machines, asserting them to be most analogically alike (e.g., a wheel-and-axle works more like a lever than it does a pulley). Criterion 2, "choose among two or more competing explanations," is particularly relevant, given that the overall goal is to make an argument about the best analogue from two possible analogues. Criterion 4, "use criteria to distinguish good from poor arguments," occurred when students undertook the discernment process about elements from simple machines as in the final two reasoning sequences analyzed. Finally, criterion 6, "persuade others" is also relevant. The need for discernment is particularly central to the process of persuasion of one's peers.

There are several limitations to this work, including that the results may not be generalizable to a larger population, those more experienced in science may be more efficient at discernment, and nonelementary education majors may perform argumentation qualitatively differently. Of course, individuals also vary in their skills necessary to engage in discernment. In addition, the activity was not the most time efficient. Repeated interactions with analogical mapping as well as instructor guidance were necessary in order for some

groups to show discernment; early attempts resulted in frustration for many students. This may not make for a good efficiency/effectiveness balance in many classrooms. Developing the analogical-mapping-based comparison activities took time. Finally, given the small sample size, instructor participation in answering questions, and the lack of data separation from other simple-machine activities, statistically significant results cannot be offered.

Although the model of activity researched here is not necessarily amenable to all science content (some science concepts are not readily comparable), similar opportunities exist for analogical comparative small-group argumentation around case comparisons, analogous laboratory experiments, learning about canonical analogies (e.g., solar-system-atomic model, electricity and water, etc.), and core ideas in science in various contexts (e.g., energy transfer, evolution, etc.).

Analogical-mapping-based comparison activities might also be relevant to research on learning progressions, which deal with the order in which content can be best learned and taught over time (National Research Council, 2007). The analogical-mapping-based comparison activities might be offered over a longer time frame, over various courses or years with ever more sophisticated models, content-explanation-analogies, or analogues. This aligns with what Bruner (1968) called the "spiral curriculum," in which content recurs again and again over time but in a slightly different form (e.g., more sophisticated models) and/or with different surrounding content.

To map and compare, as was did here, is to inherently make connections, which was the goal of the instructors. Such connection making also allows for one simple machine to lend structure (or not, as the case may be) to another simple machine, reducing the tension between problematizing and structuring to focus and channel students' attention and argumentation in the space between the analogues.

The authors wish to thank Mark Merritt, Elizabeth Larcom, and Richard Duschl for their input in this work.

REFERENCES

Bellocchi, A., & Ritchie, S. M. (2011). Investigating and theorizing discourse during analogy writing in chemistry. Journal of Research in Science Teaching, 48(7), 771–792.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). How people learn: Brain, mind, experience and school. Washington, DC: National Academy Press.

Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), Similarity and analogical reasoning (pp. 470–491). Cambridge, England: Cambridge University Press.

Brewer, S. (1996). Exemplary reasoning: Semantics, pragmatics, and the rational force of legal argument by analogy. Harvard Law Review, 109, 923–1028.

Brown, D. E., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. Instructional Science, 18, 237–261.

Bruner, J. S. (1968). Toward a theory of instruction (1st ed.). New York: John Wiley.

Clement, J. (1981). Analogy generation in scientific problem solving. In Proceedings of the Third Annual Meeting of the Cognitive Science Society, Berkeley, CA (1–12). ERIC RIE #SE-048-920.

Clement, J. (1988). Observed methods for generating analogies in scientific problem solving. Cognitive Science, 12, 263–586.

Clement, J., & Brown, D. E. (2008). Using analogies and models in instruction to deal with students' preconceptions Creative model construction in scientists and students: the role of imagery, analogy, and mental simulation. Dordrecht, The Netherlands: Springer.

Coll, R. (2006). The role of models, mental models, and analogies in chemistry teaching. In P. J. Aubusson, A. G. Harrison, & S. M. Ritchie (Eds.), Metaphor and analogy in science education (Vol. 30, pp. 65–77). Dordrecht, The Netherlands: Springer.

Diehl, V., & Reese, D. D. (2010). Elaborated metaphors support viable inferences about difficult science concepts. Educational Psychology, 30(7), 771–791.
Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. Science Education, 84, 287–312.
Duit, R. (1991). On the role of analogies and metaphors in learning science. Science Education, 75, 649–672.
Dunbar, K. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings, yet so difficult in the psychological laboratory. In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.), Analogy: Perspectives from cognitive science (pp. 313–334). Cambridge, MA: MIT Press.
Else, M. J., Clement, J., & Rea-Ramirez, M. A. (2003). Should different types of analogies be treated differently in instruction? observations from a middle-school life science curriculum. Paper presented at the Proceedings of the National Association for Research in Science Teaching, Philadelphia, USA.
Else, M. J., Clement, J., & Rea-Ramirez, M. A. (2008). Using analogies in science teaching and curriculum design: Some guidelines. In J. J. Clement & M. A. Rea-Ramirez (Eds.), Model based learning and instruction in science (pp. 215–231). Dordrecht, The Netherlands: Springer.
Fogwill, S. N. (2010). Student co-generated analogies and their influence on the development of science understanding. Sydney, Australia: University of Technology Sydney.
Gadgil, S., & Nokes, T. (2009). Analogical scaffolding in collaborative learning. Paper presented at the Proceedings of the 31st Annual Conference of the Cognitive Science Society, Austin, TX.
Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science, 7, 155–170.
Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. Journal of Educational Psychology, 95(2), 393–408.
Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. Cognitive Psychology, 12(3), 306–355.
Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. 15(1), 1–38.
Glasson, G. E. (1989). The effects of hands-on and teacher demonstration laboratory methods on science achievement in relation to reasoning ability and prior knowledge. Journal of Research in Science Teaching, 26(2), 121–131.
Glynn, S. M. (1991). Explaining science concepts: A teaching-with analogies model. In S. M. Glynn, R. H. Yeany, & B. K. Britton (Eds.), The psychology of learning science (pp. 219–240). Hillsdale, NJ: Erlbaum.
Harrison, A. G., & De Jong, O. (2005). Exploring the use of multiple analogical models when teaching and learning chemical equilibrium. Journal of Research in Science Teaching, 42(10), 1135–1159.
Hewitt, P. G. (1999). Conceptual physics. Menlo Park, CA: Scott Foresman–Addison Wesley.
Heywood, D. (2002). The place of analogies in science education. Cambridge Journal of Education, 32(2).
Jimenez-Aleixandre, M. P. (2008). Designing argumentation learning environments. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), Argumentation in science education (pp. 91–115). Dordrecht, The Netherlands: Springer.
Kurtz, K. J., & Lowenstein, J. (2007). Converging on a new role for analogy in problem solving and retrieval: When two problems are better than one. Memory and Cognition, 35(2), 334–341.
Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. Journal of the Learning Sciences, 10(4), 417–446.
Marton, F. (2006). Sameness and difference in transfer. Journal of the Learning Sciences, 15(4), 499–535.
Mascolo, M. F. (2005). Change processes in development: The concept of coactive scaffolding. New Ideas in Psychology, 23, 185–196.
Mason, L. (1996). Collaborative reasoning on self-generated analogies: conceptual growth in understanding scientific phenomena. Educational Research and Evaluation, 2(4), 309–350.
May, D. B., Hammer, D., & Roy, P. (2006). Children's analogical reasoning in a third-grade science discussion. Science Education, 90(2), 316–330.
McKenna, A., & Agogino, A. (1998). A web-based instructional module for teaching middle school students engineering design with simple machines*. Journal of Engineering Education (Washington, DC), 87(4), 437–444.
Mussweiler, T., & Epstude, K. (2009). Relatively fast! Efficiency advantages of comparative thinking. Journal of Experimental Psychology: General, 138(1), 1–21.
Nashon, S. M. (2004). The nature of analogical explanations: High school physics teachers use in Kenya. Research in Science Education 34, 475–502.
National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8 (R. A. Duschl, H. Schweingruber, & A. Schouse, Eds.). Washington, DC: National Academies Press.
National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press.
Oh, P. S. (2011). Characteristics of abductive inquiry in earth science: An undergraduate case study. Science Education, 95, 409–430.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. Journal of the Learning Sciences, 13(3), 423–451.
Podolefsky, N. S., & Finkelstein, N. D. (2007). Analogical scaffolding and the learning of abstract ideas in physics: Empirical studies. Physical Review Special Topics—Physics Education Research, 3.
Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. Cognition and Instruction, 11(3–4), 365–395.
Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. Journal of the Learning Sciences, 13(3), 273–304.
Renninger, K. A., & Granott, N. (2005). The process of scaffolding in learning and development. New Ideas in Psychology, 23, 111–114.
Roth, W.-M. (1996). Thinking with hands, eyes, and signs: Multimodal science talk in a grade 6/7 unit on simple machines. Interactive Learning Environments, 4, 170–187.
Roth, W.-M. (2001). Learning science through technological design. Journal of Research in Science Teaching, 38(7), 768–790.
Savinainen, A., Scott, P., & Viiri, J. (2005). Using a bridging representation and social interactions to foster conceptual change: Designing and evaluating an instructional sequence for Newton's third law. Science Education, 89, 175–195.
Theile, R. B., & Treagust, D. F. (1991). Using analogies in secondary chemistry teaching. Information Analyses—Guides June 1991, 22.
Tohill, J. M., & Holyoak, K. J. (2000). The impact of anxiety on analogical reasoning. Thinking and Reasoning, 6, 27–40.
Toulmin, S. E. (1958). The uses of argument (3rd ed.). Cambridge, England: Cambridge University Press.
Treagust, D. F., Harrison, A. G., & Venville, G. J. (1998). Teaching science effectively with analogies: An approach for preservice and inservice teacher education. Journal of Science Teacher Education, 9(2), 85–101.
Tucknott, J. M., & Yore, L. D. (1999). The effects of writing activities on grade 4 children's understanding of simple machines, inventions, and inventors. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Boston, MA.
von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. Journal of Research in Science Teaching, 45(1), 101–131.
Vygotsky, L. S. (1978). Mind in society (L. S. Vygotsky, M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.
Walker, K. A., & Zeidler, D. L. (2007). Promoting discourse about socioscientific issues through scaffolded inquiry. International Journal of Science Education, 29(11), 1387–1410.
Wilbers, J., & Duit, R. (2006). Post festum and heuristic analogies. In P. J. Aubusson, A. G. Harrison, & S. M. Ritchie (Eds.), Metaphor and analogy in science education (Vol. 30, pp. 37–49). Dordrecht, The Netherlands: Springer.
Wong, E. D. (1993). Understanding the generative capacity of analogies as a tool for explanation. Journal of Research in Science Teaching, 30(10), 1259–1272.
Wood, D. J., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. Journal of Child Psychology and Psychiatry, 17, 89–100.
Yerrick, R. K., Doster, E., Nugent, J. S., Parke, H. M., & Crawley, F. E. (2003). Social interaction and the use of analogy: An analysis of preservice teachers' talk during physics inquiry lessons. Journal of Research in Science Teaching, 40(5), 443–463.
Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2003). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. Research in Science Education, 32, 437–463.
Zook, K. B., & DiVesta, F. J. (1991). Instructional analogies and conceptual misrepresentations. Journal of Educational Psychology, 83(2), 246–252.

Science
Education

When Relationships Depicted Diagrammatically Conflict With Prior Knowledge: An Investigation of Students' Interpretations of Evolutionary Trees

LAURA R. NOVICK,[1] KEFYN M. CATLEY[2]
[1] *Department of Psychology & Human Development, Vanderbilt University, TN 37203-5721, USA;* [2] *Department of Biology, Western Carolina University, Cullowhee, NC 28723, USA*

Received 11 March 2012; accepted 14 November 2013
DOI 10.1002/sce.21097
Published online 13 January 2014 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Science is an important domain for investigating students' responses to information that contradicts their prior knowledge. In previous studies of this topic, this information was communicated verbally. The present research used diagrams, specifically trees (cladograms) depicting evolutionary relationships among taxa. Effects of college students' and 10th graders' prior knowledge on their ability to reason from information depicted in cladograms was evaluated in two ways: (1) By keeping the hierarchical branching structure constant while manipulating whether the taxa-targeted common misconceptions about biological classification or were unfamiliar; and (2) by keeping the targeted misconception constant while manipulating the strength of the evidence countermanding that misconception. Students demonstrated more sophisticated reasoning when (1) the taxa were unfamiliar, so they had to rely on the diagrammatic information presented rather than their incorrect prior knowledge, and (2) stronger evidence contradicting their incorrect

Correspondence to: Laura Novick; e-mail: Laura.Novick@vanderbilt.edu
Contract grant sponsor: Institute of Education Sciences, U.S. Department of Education (to Vanderbilt University).
Contract grant number: R305A080621.
The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

prior knowledge was presented. Students' challenges to correctly interpreting evolutionary trees included lower level of schooling and greater strength of the misconception. College, but not high school, students showed some ability to transfer their better reasoning with cladograms depicting relationships among unfamiliar taxa to cladograms depicting taxon relationships that contradicted their everyday conceptions. Implications for improving biology education and overcoming misconceptions are discussed. *Sci Ed* **98:**269–304, 2014

INTRODUCTION

Students often encounter information in their science classes that contradicts what they believe to be true (e.g., Chinn & Brewer, 1993, 1998; Tanner & Allen, 2005; Wandersee, Mintzes, & Novak, 1994). Thus, science is a natural domain for investigating effects of incorrect prior knowledge on learning, comprehension, and reasoning. An extensive research literature indicates that it can be difficult to persuade students to replace their scientifically incorrect, everyday conceptions with more scientifically accepted ideas (e.g., Alverson, Smith, & Readence, 1985; Chinn & Brewer, 1998; Kendeou & van den Broek, 2005; Zimmerman, 2007).

The studies on this topic have largely used verbal texts. However, diagrams, the use of which in science dates back at least to the fifteenth century (Hegarty, Carpenter, & Just, 1991), are at least as important a means of conveying information. Three types of visual representations are found in written science materials (Hegarty et al., 1991): (a) iconic diagrams, which resemble their referents (e.g., a sketch of a dragonfly, a drawing of a pulley system); (b) schematic diagrams, which depict the underlying structure of abstract concepts and rely on conventions for their use (e.g., Newman projections of molecules, evolutionary trees); and (c) charts and graphs. Schematic diagrams, in particular, are ubiquitous in the sciences because they are critical tools for developing theories, solving problems, and communicating structures, processes, and relationships (e.g., Hegarty & Stull, 2012; Kindfield, 1993/1994; Lynch, 1990; Maienschein, 1991; Novick, 2006). We investigated students' interpretations of evolutionary trees, which are a critically important type of schematic diagram in contemporary biology.

CLADOGRAMS: SCHEMATIC DIAGRAMS THAT DEPICT PORTIONS OF THE TREE OF LIFE

The most common depiction of the Tree of Life is the *cladogram*, a hierarchical diagram that shows biologists' current best-supported hypotheses about evolutionary relationships among a set of taxa in terms of nested levels of common ancestry. (A taxon is any single taxonomic group ranging from a species [e.g., *Homo sapiens*] to a higher order group [e.g., primates, mammals, amniotes, vertebrates].) The cladogram in Figure 1b, for example, shows that badgers and foxes are more closely related to each other than either is to any other taxon on that cladogram because their *most recent common ancestor* (MRCA) is not shared by any of the other taxa. Among the taxa depicted, badgers and foxes are next most closely related to mushrooms, because those three taxa share a more recent common ancestor with each other than they do with any of the other taxa. Thus, mushrooms are more closely related to badgers and foxes than to grass and geraniums. Substituting six other taxa (e.g., kinds of spiders) for those shown in Figure 1b, while keeping the branching structure (the topology) the same, yields an isomorphic cladogram depicting evolutionary relationships among the new taxa (see Figure 1a).

Another critical concept in evolutionary biology is that of a *clade*, a group comprising an MRCA and all of its descendants. From a cladistic perspective (e.g., Hennig, 1966;

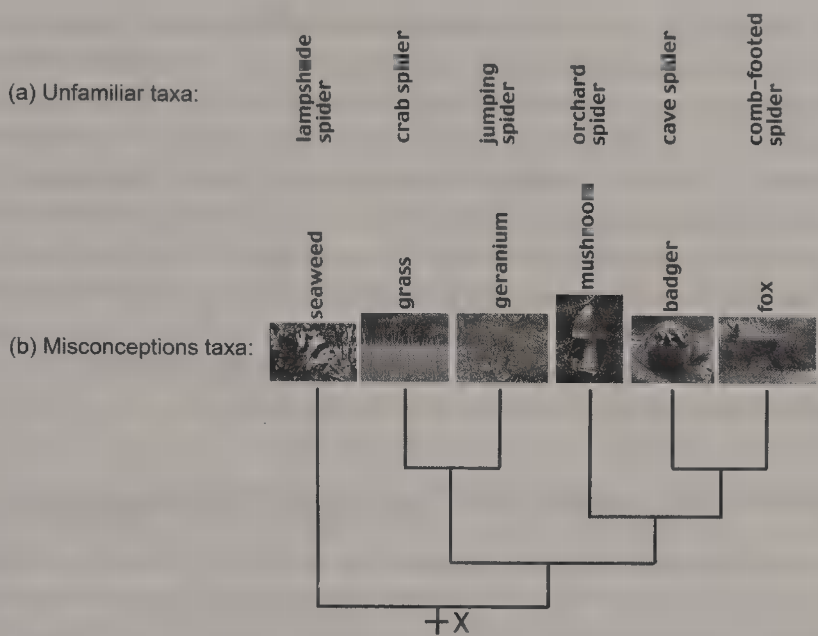


Figure 1. Structurally identical cladograms that depict relationships among (a) unfamiliar taxa and (b) familiar taxa about which students have misconceptions (identified as Structure 1 in the Appendix; used in Studies 1 and 2). The pictures of the familiar taxa were presented in color in the materials students received.

Thanukos, 2009), only clades are valid biological groups. Groups that exclude one or more descendants of the MRCA (paraphyletic groups) have no basis in evolutionary history and thus are noninformative. This is why humans are classified as primates—they are one of the descendants of this group's MRCA. Similarly, a large body of evidence from evolutionary biology in the form of molecular, morphological, and behavioral data accumulated over the past 20 years or so unequivocally indicates that birds, like all dinosaurs, are reptiles (e.g., Freeman, 2011; Lee, Reeder, Slowinski, & Lawson, 2004; Reece et al., 2011; Thanukos, 2009). Because we are interested in science education and students' understanding of science, when we state that birds are reptiles in this article we are referring to scientific classification. Of course, students may harbor misconceptions about biological classification just as they do about topics in physics and chemistry.

Cladograms are key inferential tools in biology that have yielded considerable benefits to humankind with respect to, for example, health, agriculture, and biotechnology (e.g., American Museum of Natural History [AMNH], 2002; Freeman, 2011; Futuyma, 2004; Yates, Salazar-Bravo, & Dragoo, 2004). The ability to understand and reason with the information depicted in cladograms is referred to as *tree thinking*. A number of scholars (e.g., Baum, Smith, & Donovan, 2005; Catley, 2006; Gilbert, 2003; Goldsmith, 2003; O'Hara, 1988) have argued for the need to include tree thinking in evolution curricula, which has led to the appearance of cladograms in introductory biology textbooks at both the high school and college levels (Catley & Novick, 2008).

THEORETICAL BACKGROUND

Diagrammatic literacy underpins conceptual development in science (e.g., Bowen & Roth, 2002; Maienschein, 1991). In contemporary biology, understanding cladograms is an essential skill (e.g., Novick & Catley, 2013; Thanukos, 2009). For example, consider Bio.3.5.2 from the North Carolina Standard Course of Study for high school biology: "Analyze the classification of organisms according to their evolutionary relationships

(including dichotomous keys and phylogenetic trees)" (retrieved August 6, 2013, from http://www.dpi.state.nc.us/acre/standards/new-standards/). But, despite the importance of diagrammatic thinking in science, little is known about how students reconcile their pre-existing misconceptions with contradictory information presented in diagrams generally or in cladograms specifically. Such knowledge is critical for educators to design effective curricula to teach tree thinking and to promote conceptual change in the many areas of science that require diagrammatic thinking. We investigated the influence of college and high school students' misconceptions on their ability to reason with the fundamental tree-thinking concepts of most recent common ancestry and a clade. Our research draws on theoretical perspectives from the psychology of diagrammatic reasoning and from science education research on students' responses to anomalous data.

On the Superiority of Diagrammatic Over Textual Representations

A large body of research has documented the benefits of visual over verbal representations for learning, reasoning, and problem solving (e.g., Ainsworth & Loizou, 2003; Hegarty & Just, 1993; Kindfield, 1993/1994; Rotbain, Marbach-Ad, & Stavy, 2006; Sweller, Chandler, Tierney, & Cooper, 1990). Schematic diagrams, in particular, are important tools for thinking (e.g., Dufour-Janvier, Bednarz, & Belanger, 1987; Kindfield, 1993/1994; Larkin & Simon, 1987) because they (a) simplify complex situations (Lynch, 1990; Winn, 1989), (b) make abstract concepts more concrete (Winn, 1989), and (c) substitute easier perceptual inferences for more cognitively demanding search processes and verbal deductive inferences (Barwise & Etchemendy, 1991; Larkin & Simon, 1987). For example, compare the following verbal representation of evolutionary relationships to the equivalent diagrammatic representation shown in Figure 1b: (seaweed + ((grass + geranium) + (mushroom + (badger + fox)))). Moreover, consider the difference in interpretability of such verbal and diagrammatic representations as the number of taxa increases.

Because what is immediately apparent visually has a profound effect on people's understanding, when students' prior knowledge conflicts with accepted scientific interpretations, it may be beneficial to present scientific information diagrammatically. That is, a diagrammatic depiction of the current scientifically accepted relationships may be powerful enough to overcome students' misconceptions. If this is the case, students should respond similarly to tree-thinking questions about a cladogram that depicts relational information that conflicts with their prior knowledge as when no such conflict is present. With respect to Figure 1, then, students would be just as likely to group mushroom with badger and fox rather than grass and geranium as they are to group orchard spider with the cave and comb-footed spiders rather than with the crab and jumping spiders.

Responding to Anomalous Data in Science

Students, however, can be tenacious in holding onto misconceptions in the face of strong evidence to the contrary. To understand how students respond to diagrammatic depictions that contradict their prior knowledge, we adopted Chinn and Brewer's (1993, 1998) theoretical framework for interpreting people's responses to anomalous data, i.e., data that contradict their current theory of the physical world. A key outcome of their research was a taxonomy of eight types of responses to anomalous data: (a) ignore the data, (b) reject the data, (c) express uncertainty as to the believability of the data, (d) exclude the data from the domain of the theory, (e) hold the data in abeyance, (f) reinterpret the data while retaining the existing theory, (g) make peripheral changes to the existing theory given one's reinterpretation of the data, and (h) accept the data and adopt a new theoretical

explanation. Chinn and Brewer (1993, 1998) found evidence to support this taxonomy from both new data collection involving college students and from secondary analyses of existing empirical and historical data concerning responses of primary and secondary students and established scientists.

In their 1998 study, Chinn and Brewer evaluated college students' responses to written descriptions of theories and associated anomalous data. Briefly, students first read a lengthy description of an initial theory (e.g., the volcanic eruption theory of the mass extinctions at the end of the Cretaceous Period) and then read about some data that were anomalous given that theory (e.g., the eruptions were not strong enough to have had this effect). Belief ratings collected after students read the initial theory confirmed that nearly all students strongly believed the theory. After reading about the anomalous data, students rated the extent to which they believed the data, provided a written explanation for their rating, rated the consistency of the data with the theory, and provided a written explanation for that rating. Finally, students rated their belief in the initial theory again and provided a written explanation. Few students (4.8%) changed their theory given the anomalous data. Instead, students provided a variety of reasons for discounting the anomalous data, which Chinn and Brewer were able to classify into the eight categories mentioned above. The most common response was to reject the data (34.5% of responses). Other common explanations involved reinterpretations of the data that enabled students to retain their current theory (24.3%) or expressions of uncertainty about the believability of the data (17.5%). In contrast, only 1.7% of responses suggested peripheral theory change. Each of the remaining response types accounted for fewer than 10% of responses. Chinn and Brewer attributed the low frequency of ignoring the data (8.5%) to the fact that students were required by the experimental procedure to evaluate the data both numerically and by providing a written explanation.

Applying Chinn and Brewer's (1993, 1998) Theory to Tree Thinking

Chinn and Brewer (1993) framed their research question as one of understanding how students respond to scientific information that contradicts their current theories of the world. In discussing their work, they distinguished three important concepts—beliefs, knowledge, and theories. They defined *beliefs* as specific pieces of knowledge within a student's knowledge base. *Knowledge* was used to refer to the student's total set of beliefs. A *theory* was defined as a collection of beliefs that had explanatory force for the student.

We examined whether Chinn and Brewer's (1993, 1998) theoretical perspective on responses to anomalous data is applicable to a situation that differs from the ones they investigated in several respects: (a) The initial theories Chinn and Brewer's (1998) students had were those they had just learned in the study. Our students' prior knowledge differed in three ways. First, that knowledge was acquired prior to, and potentially long before, participating in our study. Second, it was scientifically inaccurate and thus constitutes a misconception (e.g., that mushrooms are more closely related to plants than to animals). Third, these misconceptions would be better characterized as beliefs rather than theories. (b) The contradictory information Chinn and Brewer (1998) presented to students was in the form of anomalous data. In our study, the contradictory information (e.g., that mushrooms actually are more closely related to animals than to plants) was presented in the form of cladograms. Cladograms represent biologists' current best-supported hypotheses of evolutionary relationships among the indicated taxa, based on large, complex data sets that students rarely see. (c) In Chinn and Brewer's (1998) study, both the initial information (i.e., the theory) and the contradictory information (the anomalous data) were presented verbally (as written text). In our study, students retrieved the initial information from

memory and received the contradictory information in a diagrammatic format (specifically, a cladogram).

From a student's perspective, the difference between scientific data that contradict a currently accepted theory and scientific information that contradicts a current belief seems minimal. In both cases, the student encounters new information that conflicts with what he/she already believes to be true and has to figure out how to respond to that contradiction. We see no reason to think that Chinn and Brewer's (1993, 1998) taxonomy would be restricted to situations in which the conflict is between data and theory. It is, however, an open question whether students would offer the same types of responses to the conflicting information when that information is presented diagrammatically rather than verbally. As noted earlier, diagrams may provide a more powerful way to present information that contradicts students' prior beliefs.

FRAMEWORK FOR OUR STUDIES

In many areas of science, critical information is presented using diagrams rather than words (e.g., Bowen & Roth, 2002; McKim, 1980). It is incumbent on educators, therefore, to understand how students interpret and reason with information in this format, including how their interpretations are affected by their prior beliefs regarding the information presented.

Why Study Tree Thinking?

Evolutionary biology is an ideal, yet untested, domain for investigating students' responses to contradictory information presented diagrammatically. In cladograms, the medium of choice for depicting historical evolutionary relationships, the nested branching pattern provides abstract information about relationships that is applied to specific taxa labeling the terminal branches—e.g., spiders in Figure 1a and plants, fungi, and animals in Figure 1b. Tree-thinking questions are questions about the branching pattern.

This feature of cladograms enables a powerful test of the effect of prior knowledge because the same branching pattern can be applied to taxa about which students do or do not have conflicting prior knowledge. For example, the relationships shown in Figure 1b, which are consistent with contemporary scientific research, conflict with students' prior knowledge because people generally think mushrooms are plants (Goldberg & Thompson-Schill, 2009; Hampton, 1988). In contrast, the isomorphic relationships among types of spiders shown in Figure 1a do not conflict with students' prior knowledge because students are unlikely to know anything about spider relationships. Because the cladograms in Figures 1a and 1b depict the same topology, from the perspective of evolutionary biology the identities of the taxa are irrelevant to answering tree-thinking questions about those cladograms: For example, whatever taxon is in the grass/crab spider position is more closely related to whatever taxon is in the geranium/jumping spider position than it is to any other taxon on the cladogram. Nevertheless, students' responses may be affected by their prior knowledge of the taxa. Nehm and Ha (2011) and Opfer, Nehm, and Ha (2012) found that the content of written scenarios affected students' responses to questions about natural selection. Our manipulation of the relation between students' prior knowledge and the information presented is novel as is our use of a diagrammatic presentation format.

There is a growing theoretical, descriptive, and empirical literature on tree thinking that has considered both the nature of students' successes and failures at tree thinking and the effects of instruction on improving tree-thinking skills (e.g., Baum et al., 2005; Gregory, 2008; Halverson, Pires, & Abell, 2011; Meir, Perry, Herron, and Kingsolver, 2007;

Novick & Catley, 2013; Phillips, Novick, Catley, & Funk, 2012; Sandvik 2008). Although some earlier articles considered students' misconceptions, these were misconceptions in how to interpret cladogram structure, regardless of the specific taxa depicted. In contrast, the focus of our studies is on misconceptions about the relationships among particular taxa. Because we compared students' responses to cladograms with the same branching structure but different taxa (see Figure 1), any difficulties students have in interpreting our branching patterns would apply equally to the taxa about which they do versus do not have misconceptions concerning their evolutionary relationships.

Effects of Prior Knowledge While Keeping Branching Structure Constant

Basic Predictions. Studies 1 and 2 examined the effects on tree thinking in college and high school students, respectively, of having conflicting prior knowledge by assigning different sets of taxa to the branch tips for a cladogram illustrating a particular nested structure. Figure 1 illustrates one of three such pairs of cladograms. For each pair, one set of taxa targeted a documented misconception about relationships among familiar living things (e.g., Figure 1b). The other set of taxa was relatively unfamiliar to students (e.g., Figure 1a). Opfer et al. (2012) manipulated the familiarity of taxa and characters used in written scenarios testing students' understanding of natural selection. They found that students mentioned more key concepts of natural selection for the scenario involving a familiar animal and character (a snail that is poisonous) than for the scenarios involving an unfamiliar animal and character or familiar or unfamiliar plants.

Our manipulation of familiarity was different from that of Opfer et al. (2012) because we selected sets of familiar taxa about which the prior literature indicates that students possess misconceptions concerning the relationships among those taxa. Accordingly, we expected either of two patterns of results, both of which differ from what Opfer et al. found. If the diagrammatic presentation format is strong enough to counteract students' misconceptions, students' responses to our tree-thinking questions assessing their understanding of and ability to apply the concepts of most recent common ancestry and clades should be equally accurate for the familiar and unfamiliar taxa because the two cladograms depicted exactly the same branching structure. If the misconceptions carry more weight, however, students should be more accurate at answering the questions about the unfamiliar than the familiar taxa because their prior knowledge about the familiar taxa provides a competing basis for responding. In addition, they should be more likely to refer to their prior knowledge for the cladograms that depict relationships among familiar taxa. We will henceforth refer to the two sets of taxa as unfamiliar taxa and misconceptions taxa.

After answering the tree-thinking questions, students gave written explanations for those responses, which were evaluated for whether they fit Chinn and Brewer's (1993, 1998) categories of responses to anomalous data. We predicted that Chinn and Brewer's taxonomy would capture students' responses to the conflict between cladogram structure and their misconceptions about the indicated taxa. We also predicted that responses fitting the category of ignoring the anomalous data would be more prevalent in our studies than in Chinn and Brewer's (1998) study because our students were not specifically told to evaluate the contradictory information.

Finally, Chinn and Brewer (1998) suggested that younger students might produce a narrower range of responses to anomalous data than undergraduates. We tested this hypothesis by comparing students' responses across Studies 1 and 2. We further predicted that the high school students would produce proportionally more responses at lower levels in Chinn and Brewer's (1993, 1998) taxonomy—e.g., ignoring or rejecting the contradictory information.

We also expected 10th graders to show a larger decrement in accuracy when the cladograms depicted relationships among misconceptions taxa as opposed to unfamiliar taxa.

Order of Presenting Cladograms With Unfamiliar Versus Misconceptions Taxa. The cladograms with misconceptions and unfamiliar taxa were presented in counterbalanced blocks. The within-subjects manipulation of taxon type enabled a strong test of the effect of this factor on students' reasoning. In addition, by counterbalancing the order of presenting the two types of taxa, we could investigate whether exposure to materials of one type might positively or negatively affect reasoning with materials of the other type, which is useful to know for designing instruction.

The simplest result is that students respond similarly to cladograms depicting relationships among a given type of taxa regardless of presentation order. However, the conceptual change literature indicates that students can sometimes apply their understanding of an instructional example to a subsequent case about which they previously had misconceptions (e.g., Stavy, 1991; see Clement, 2008, for a review). Thus, we might find that responding to the cladograms with unfamiliar taxa first helps students appropriately attend to cladogram structure, which might yield benefits for responding to the cladograms with misconceptions taxa subsequently. At the same time, there is evidence from the problem-solving literature of college students who were relative novices in the domain of mathematics incorrectly applying a solution procedure from an initial problem to a subsequent problem that seemed similar but was actually different (Novick, 1988). In the present studies, this might appear as inappropriate carryover of a strategy to focus on prior knowledge rather than cladogram structure from the misconceptions taxa to the unfamiliar taxa.

Effects of Branching Structure on Reasoning About a Particular Misconception

In Studies 1 and 2, we kept the cladogram structure constant and varied the taxa to which that structure was applied. In Study 3, we adopted the complementary strategy of varying the cladogram structure relevant to a single misconception. The misconception we targeted was that birds are not reptiles. Across two cladograms, we manipulated the strength of the evidence supporting the scientifically accepted conclusion that birds in fact are reptiles. Scientists are sensitive to the strength of the evidence suggesting that a new theory should be adopted in place of the current theory (e.g., Chinn & Brewer, 1993). Accordingly, we investigated whether college students are sensitive to the strength of the evidence, provided by a cladogram's structure, countermanding their prior knowledge concerning the classification of birds. Given that tree thinking is a critical twenty-first century skill and that cladograms depicting relationships among familiar taxa will necessarily sometimes present information that conflicts with students' prior knowledge (e.g., showing that birds are reptiles), it is important to understand whether students are sensitive to how strongly the relationships depicted in the cladogram contradict their prior beliefs about taxonomic relationships.

EXPERIMENTAL PROTOCOLS FOR ASSESSING PRIOR KNOWLEDGE

There are two common methodologies for investigating effects of prior knowledge. One method requires pretesting a large group of students concerning their prior knowledge, selecting from among that group those who are known to possess the misconception(s) of interest, and then administering the experimental task to that subset of students. Typically,

the experimental task must be administered at some later point in time so that students do not connect the experimental task with the prior knowledge assessment. Logistically, this method was not available to us.

Therefore, we employed an alternative method that requires reviewing existing sources of information about students' knowledge in the relevant area(s) and selecting misconceptions to study that are relatively common in the population at large. We relied on two sources for such information: The published literature and the second author's professional experience as an instructor of the biodiversity course from the introductory biology sequence for science majors at his university. The students in that class have more knowledge of biology than do the college students in the present studies, nearly all of whom have *not* taken introductory biology for majors; yet, they still exhibit the same misconceptions as the general public about relationships among common taxa. We provide evidence on people's misconceptions at the relevant points in the presentations of the studies. This method of investigating effects of students' misconceptions rests on the (reasonable) assumption that students who receive the experimental task have the same misconceptions as most people and that students in different conditions are equivalent in this regard due to random assignment to conditions.

STUDY 1

Method

Subjects. The subjects were 70 undergraduates (34 females, 34 males, 2 undisclosed sex) at a highly selective, private, Research I institution in the southeastern United States who were recruited from the paid subject pool coordinated by the psychology department and paid $15 for their participation. Their average year in school was 2.59 (2 = sophomore, 3 = junior). At the end of the study, the students were asked whether they had taken any of 13 primarily organismal biology classes at their university (or equivalent courses elsewhere). On average, they reported having taken just over half of a one-semester class ($M = 0.59$, $SD = 0.85$, range of 0–3). We did not ask students about their prior instruction regarding cladograms and tree thinking. However, their accuracy in this study for the cladograms with unfamiliar taxa was quite high ($M = 0.84$), indicating that our tree-thinking task was well within their capability.

We also did not ask students to identify their race, ethnicity, or state of origin. The undergraduate population of the university is 70% White, 9% Black, 9% Hispanic, 6% Asian or Hawaiian/Pacific Islander, 1% Native American, and 5% mixed race. Roughly half of the students come from a southern state. New York and Illinois also send relatively large contingents of students ($\geq$5% each), followed by Maryland, Ohio, New Jersey, and other countries ($\geq$3% each).

Design. There were two independent variables. Whether the cladograms involved unfamiliar or misconceptions taxa was manipulated within subjects. The two types of cladograms were blocked. Block order was manipulated between subjects, with students randomly assigned to receive the misconceptions taxa cladograms either first ($n = 34$) or second ($n = 36$).

Materials

Overview. We created nine cladograms based on contemporary scientific research in evolutionary biology. One cladogram probed how students respond to evidence that birds are reptiles. The nature of this cladogram and students' responses to it will be discussed

as Study 3a. The remaining eight cladograms belonged to four matched sets. The two cladograms in each set presented identical nested structures (i.e., cladogram topologies) but differed in whether they depicted relationships among unfamiliar or misconceptions taxa. Each cladogram was printed near the top of an 8.5 × 11 inch piece of paper.

The misconceptions taxa were indicated by both a name and a color picture, whereas the unfamiliar taxa were presented as names only. Several factors guided this choice of presentation styles. On the one hand, two of the three sets of unfamiliar taxa (kinds of yeast and streptococcus bacteria) could not be accompanied by distinguishable pictures. Although we probably could have found distinguishable pictures of the spiders, it made more sense to us to adopt the same presentation style for all the cladograms with unfamiliar taxa. On the other hand, cladograms in textbooks often include pictures of the taxa labeling the branch tips when those taxa are expected to be familiar to students. Pictures draw students' attention and may help them retrieve previously learned information about those taxa from memory. Returning for a moment to the cladogram with the unfamiliar spiders, providing pictures of the spiders could not cue prior knowledge if such knowledge does not exist. College students cannot identify species and genera of spiders without having taken a class on spiders; none of the students in this study had taken such a class because it is not offered at the university from which they were recruited.

Following Chinn and Brewer's (1993, p. 24) suggestion that information that contradicts students' prior knowledge should "pass the test of credibility," we printed the following statement immediately above each cladogram: "The students in a basic biology class are learning about evolutionary relationships among taxa. According to biologists, the following diagram provides this information about the indicated taxa, which are various *entities*. [*entities* was replaced on each page with a category label appropriate for the taxa depicted in that particular cladogram: e.g., *spiders* for the cladogram in Figure 1a and *living things* for the cladogram in Figure 1b.] The students understand that all of the taxa shown in this diagram share a common ancestor marked by the X. Use this diagram to answer the questions on this page." The X was presented at the root of the cladogram, as illustrated in Figure 1. This introduction to each cladogram also served to cue students that we were requesting answers based on current scientific classification.

Assessing Tree Thinking. Novick and Catley (2013) introduced a taxonomy of five core tree-thinking skills. We chose two fundamental skills to provide a summary measure of tree thinking in this study. Structurally identical questions were asked about each cladogram in a matched pair.

The first question asked students to recognize a valid biological group (i.e., a clade) based on the evidence provided. The question began as follows: "The following students disagree about what subsets of these taxa are valid biological groups. Which student's subset best reflects evolutionary evidence?" Three subsets proposed by three named students followed. The two incorrect subsets comprised taxa that we expected subjects to view as potentially attractive alternatives to the correct response given data on misconceptions about the familiar taxa, discussed later. The answer alternatives for these questions are given in the Appendix. After selecting the subset they thought comprised a valid group, students were asked to provide a written explanation.

For the second question, students were told that one taxon on the cladogram has a certain character (see the Appendix) and were asked "which other taxa (could be one or more) is/are most likely to share this character?" The best answer to this inference question is to choose the taxon or taxa that is/are in the smallest clade that includes the reference taxon named in the question. Students were asked to provide a written explanation for their answer to this question as well.

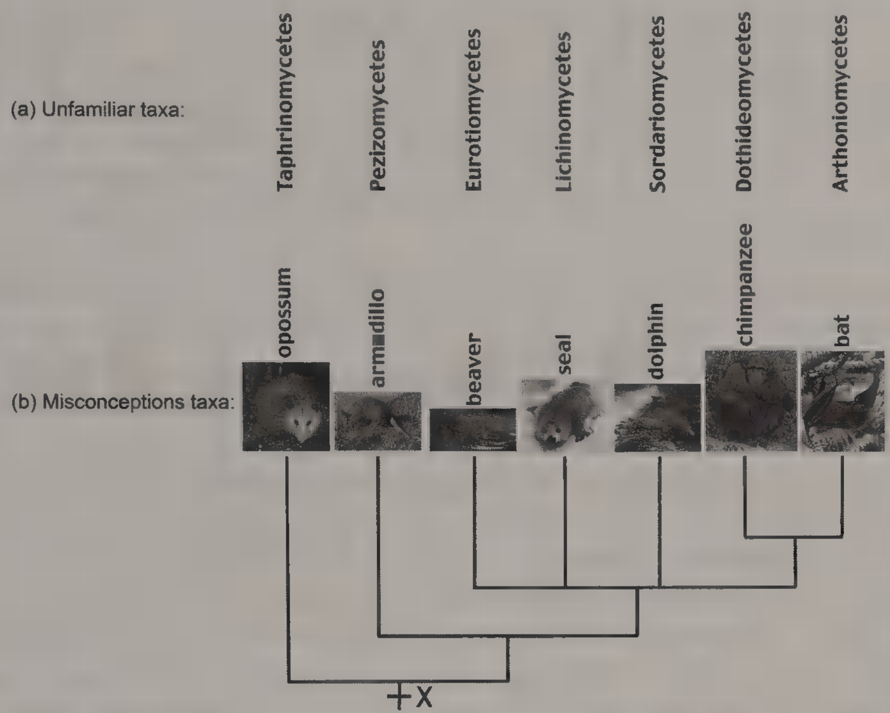


Figure 2. Structurally identical cladograms that depict relationships among (a) unfamiliar taxa and (b) familiar taxa about which students have misconceptions (identified as Structure 2 in the Appendix; used in Studies 1 and 2). The pictures of the familiar taxa were presented in color in the materials students received.

Taxon Familiarity. Obviously, familiarity of taxa falls on a continuum. Following Nehm, Beggrow, Opfer, and Ha (2012), we used Google Books ngram viewer (Michel et al., 2011; http://ngrams.googlelabs.com) to validate that the taxa we designated as familiar were in fact more familiar than those we designated as unfamiliar and that the unfamiliar taxa were unfamiliar in an absolute sense.[1] The familiar taxa in Figures 1–3 had estimated frequencies of 0.00050, 0.00017, and 0.00020, respectively. These numbers are average percentages of all words or two-word phrases in the corpus, as appropriate for the particular taxon name. The unfamiliar taxa in Figure 1 had a mean frequency that was nonzero, although it looks like zero when rounded to five decimal places. The unfamiliar taxa in Figures 2 and 3 all had frequencies of zero. Thus, the designated familiar taxa for the matched set illustrated in Figure 1 are 1,627 times more familiar than the unfamiliar taxa in that set. We could not compute the comparable comparison for the matched sets of taxa in Figures 2 and 3 because that would involve dividing by zero. To provide a lower bound estimate of the disparity in frequency, we used the unfamiliar taxa in Figure 1 as a basis for comparison. With this comparison, the designated familiar taxa in Figures 2 and 3 are 536 and 639 times more familiar than the unfamiliar taxa, respectively.

Misconceptions. One matched pair of cladograms is shown in Figure 1. As noted earlier, the familiar taxa target people's misconception that mushrooms are plants. In a categorization study involving college students, Hampton (1988) found that 70% said that mushrooms are plants. Consistent with this folk-biological classification, Goldberg and Thompson-Schill (2009) included fungi as stimuli in their category of plants. Contradicting this misconception, the correct answer to the clade question for the cladogram in Figure 1b

[1] We used the American English corpus for the years 1996–2008, with smoothing set to 10 to help estimate the average. This range of dates was chosen because it includes references that are contemporary for the students in our studies.

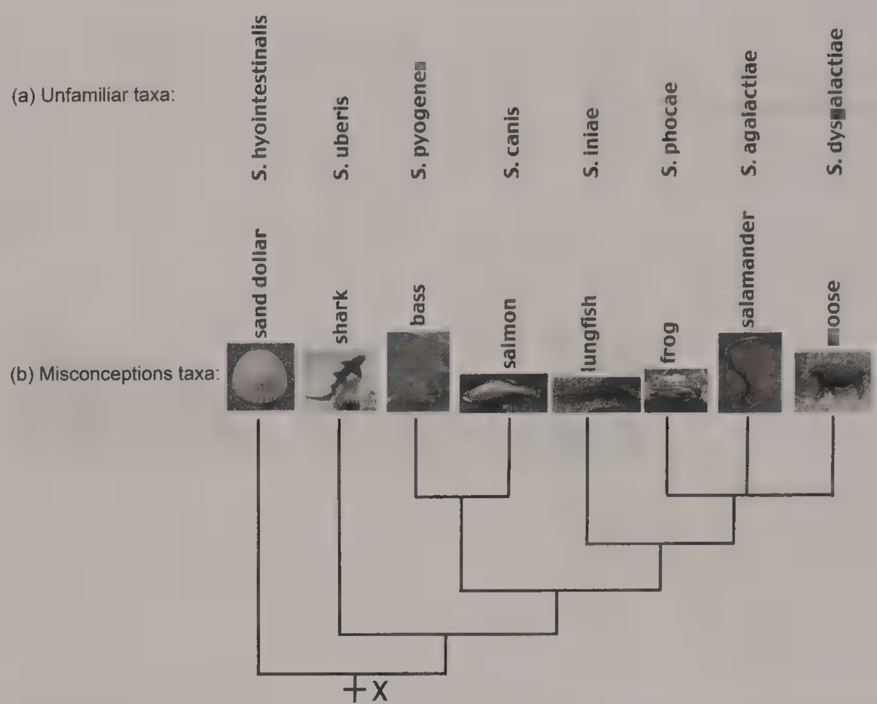


Figure 3. Structurally identical cladograms that depict relationships among (a) unfamiliar taxa and (b) familiar taxa about which students have misconceptions (identified as Structure 3 in the Appendix; used in Studies 1 and 2). The pictures of the familiar taxa were presented in color in the materials students received.

is *mushroom* + *badger* + *fox* because, of the three choices, only these taxa comprise all the descendants of a MRCA. (The corresponding correct answer for the cladogram with unfamiliar taxa is *orchard spider* + *cave spider* + *comb-footed spider*.) The two incorrect subsets of taxa included, given the targeted misconception, alternative sets of plants: *seaweed* + *grass* + *geranium* and *grass* + *geranium* + *mushroom* (*lampshade spider* + *crab spider* + *jumping spider* and *crab spider* + *jumping spider* + *orchard spider*, respectively, for the matched cladogram). The three sets of taxa were presented in a different order for the two matched cladograms and were proposed by hypothetical students with different names. The inference question for the mushroom cladogram asked which taxa are most likely to share a character of mushrooms. The correct answer is badger and fox.

A second matched pair of cladograms is shown in Figure 2. The familiar taxa target the misconception that habitat is an important basis for determining which taxa are closely related (e.g., Morabito, Catley, & Novick, 2010). The correct answer to the clade question for Figure 2b is *beaver* + *seal* + *dolphin* + *chimpanzee* + *bat*, a set of taxa found in three separate habitats: water, land, and air. The incorrect alternative we expected to be most appealing included the three aquatic mammals: *beaver* + *seal* + *dolphin*. The other incorrect alternative was *dolphin* + *chimpanzee*, which are both intelligent mammals. In one study, Osherson, Smith, Wilkie, and López (1990) asked college students to evaluate the similarity of dolphins to each of nine other mammals, including seals and chimpanzees. Dolphins were viewed as most similar to seals (score of 0.92 on a 0–1 scale), next most similar to chimpanzees (score of 0.50), and not similar to each of the other seven mammals ($M = 0.27$). In a second study, Osherson et al. found that students thought information about dolphins and seals was highly unlikely to generalize to mammals as a whole, supporting the idea that aquatic mammals are viewed as their own separate category. The inference question for the cladogram in Figure 2b asked which taxa are most likely to share a character possessed by chimpanzees. The correct answer is bat.

A third matched pair of cladograms is shown in Figure 3. The familiar taxa target the misconception that the most inclusive meaningful groups of animals are those at the folk-biological rank of life form, such as land mammals, birds, "reptiles," amphibians, fish, and insects (e.g., Atran, 1998; Berlin, Breedlove, & Raven, 1973). The correct answer to the clade question for Figure 3b is *lungfish* + *frog* + *salamander* + *moose*, which requires grouping taxa comprising three distinct life forms. The two incorrect alternatives each comprise taxa from a single life form: fish (*shark* + *bass* + *salmon* + *lungfish*) or amphibians (*frog* + *salamander*). The inference question for this cladogram asked which taxa are most likely to share a character possessed by lungfish. The correct answer is frogs, salamanders, and moose.

The fourth matched pair of cladograms (see the Appendix) will not be discussed. For the cladograms shown in Figures 1–3, accuracy for the clade questions was at least 0.10 higher for the unfamiliar than the misconceptions taxa. For the misconceptions taxa, the highest accuracy score was 0.77. For the fourth pair of cladograms, however, accuracy was at ceiling for the misconceptions taxa (0.97) and somewhat lower for the unfamiliar taxa (0.86). Two factors appear to have led to this aberrant pattern. First, the topology of the cladogram (a group of three—the correct answer—plus a group of four) provided a strong perceptual cue to the correct answer, as the mean accuracy across both versions of this cladogram was 0.91, compared with means of 0.51–0.83 for the other three structures.[2] Second, quite a few students appeared to confuse the stone and Scots pines in the cladogram with unfamiliar taxa, presumably because they are one-syllable words that begin with *s*, as they selected the answer that included the Scots pine (incorrect) rather than the stone pine (correct) in addition to the Turkish and Aleppo pines. Given these problems with this matched pair, we did not use it in Study 2. Because a key goal of our research was to compare the effects of misconceptions across college and high school students, the analyses of the Study 1 data also included only the matched pairs of cladograms shown in Figures 1–3.

Problem Booklets. The nine cladograms were presented in two blocks: the five with misconceptions taxa (including the one discussed in Study 3a) followed by the four with unfamiliar taxa or vice versa. Within each block, the cladograms were ordered so that those that are similar (e.g., the two cladograms in a matched pair, two cladograms with similar structures) did not appear consecutively and so that a particular position for the correct answer to the first question was not used for more than two cladograms in a row. Two different orders of the cladogram pages were used for each block order.

Procedure. Students completed three booklets addressing distinct conceptual issues in a single session lasting approximately 50–75 minutes. The first booklet presented four cladograms involving familiar taxa, about which we asked the following kinds of questions: (a) What does this diagram show about the evolution of taxon X, (b) which taxa are most closely related, (c) which taxon is most highly evolved, and (d) mark all the clades in the diagram (the definition of a clade was provided). The second booklet presented eight cladograms in a diagonal (ladder) format, and students had to redraw them in the rectangular (tree) format used in the present study (and the first booklet). These cladograms used unfamiliar taxa with Latin names, so students had to attend solely to the structure of the cladograms to complete the task. Neither of these two sets of materials manipulated the presence of misconceptions. The results of these studies are reported elsewhere. The third booklet contained the materials for the present study (and Study 3a). Finally, students

[2]The role of perceptual grouping on students' interpretations of cladograms is an important question that is beyond the scope of the present study.

completed a questionnaire at the end of the session that asked for background information such as year in school and biology courses taken. Although some sessions included multiple students, students completed the booklets individually and without consulting outside resources.

Results and Discussion

Accuracy. For the clade questions, students received a score of 1 for selecting the correct answer and a score of 0 otherwise. For the inference questions, they received a score of 1 for listing all the taxa in the smallest clade that included the reference taxon named in the question, a score of 0.5 for listing a subset of those taxa, and a score of 0 otherwise. These scores were averaged across the cladograms with unfamiliar versus misconceptions taxa to yield a composite tree-thinking accuracy score for each type of taxa. These scores were submitted to a 2 (type of taxa; within) × 2 (block order: unfamiliar/misconceptions vs. misconceptions/unfamiliar; between) mixed analysis of variance (ANOVA).[3] An alpha level of .05 was the criterion for statistical significance. Effect size is reported as η_p^2. Following Cohen's (1988) guidelines for proportion of variance accounted for, 0.01, 0.09, and 0.25 are the minimum values taken to indicate, respectively, a small, medium, and large effect.

There was a significant main effect of type of taxa, $F(1, 68) = 20.29$, $p < .001$, $MSE = 0.01$, $\eta_p^2 = 0.23$, with students doing worse when common misconceptions about the depicted taxa were cued ($M = 0.75$) than when no relevant prior knowledge was cued ($M = 0.84$). In a follow-up analysis, we examined the first block of data only, which yields a between-subjects design with students randomly assigned to receive either the misconceptions or unfamiliar taxa. The one-factor ANOVA yielded a significant main effect of type of taxa, $F(1, 68) = 21.34$, $p < .001$, $MSE = 0.04$, $\eta_p^2 = 0.24$. Students who received the unfamiliar taxa had much higher accuracy scores than did students who received the misconceptions taxa, with means of 0.89 and 0.66, respectively. The results of both analyses indicate that the diagrammatic depiction of information contradicting inaccurate prior knowledge was not strong enough by itself to overcome these misconceptions. The observed pattern of higher tree-thinking scores for unfamiliar than familiar taxa is opposite of what Opfer et al. (2012) found for students reasoning about natural selection. Taken together, the results of our study and Opfer et al.'s study indicate that what is important is not familiarity per se but the accuracy of students' prior knowledge about the familiar taxa.

There was also a significant main effect of block order, $F(1, 68) = 10.79$, $p < .01$, $MSE = 0.06$, $\eta_p^2 = 0.14$: Students who received the cladograms with unfamiliar taxa first ($M = 0.86$) did better than those who received the cladograms with misconceptions taxa first ($M = 0.72$). These two factors did not interact, $F(1, 68) = 2.74$, $p > .10$, $\eta_p^2 = 0.04$. The block order effect suggests the presence of both positive and negative carry-over effects, supporting the idea that students adopted a strategy based on experience with the first set of cladograms, privileging either cladogram structure (when the unfamiliar taxa came first) or prior knowledge (when the misconceptions taxa came first), which they carried forward to the second set of cladograms. Carrying over a focus on the cladogram structure to the cladograms with misconceptions taxa would improve accuracy for those cladograms, thereby increasing accuracy for the full set of cladograms overall. In contrast, carrying over a strategy to search memory for information on which to base one's response to the

[3]Preliminary analyses indicated no differences as a function of the order in which the problems within each block were collated.

cladograms with unfamiliar taxa could reduce accuracy for those cladograms as the absence of relevant information in memory might increase the likelihood of guessing. We consider implications of these results with respect to teaching tree thinking in the general discussion.

Prior Knowledge Explanations. Students were told to respond to the questions based on the evolutionary evidence shown in the cladograms. We were most interested in the explanations they gave when they ignored this instruction and referred to their prior knowledge. Examples of cladogram-based, although not necessarily correct, explanations include the following: (a) "these follow & preceede [sic] the mushroom" (Figure 1b), (b) "all and only these 5 share a certain common ancestor" (Figure 2b), (c) "they are from the same branch" (Figure 2b), and (d) "Frog, Salamander, & moose all evolve from the lungfish" (Figure 3b). Examples of prior knowledge explanations include "grows on land & plant" (Figure 1b) and "both amphibians" (Figure 3b).

Frequency of Occurrence. Students' explanations were first coded as referring to prior knowledge or not. This was done independently by two coders, who agreed on the coding of 98% of the responses for the college and high school students in Studies 1 and 2. The agreement rate was nearly identical for the two groups of subjects. Disagreements were resolved by discussion.

The college students in this study mostly followed the instructions and gave explanations based on the information in the cladograms. Although prior knowledge explanations were not common, as predicted they occurred only for the cladograms with misconceptions taxa: $M = 8.6\%$ of explanations for those three cladograms across the clade and inference questions, compared with 0% for the matched cladograms with unfamiliar taxa. The prior knowledge explanations were given by 18 students (26% of the total), who gave one to five such explanations each ($M = 2.06$). As expected, these explanations primarily supported incorrect answers: For each of these students, we computed mean accuracy for the cladograms with misconceptions taxa as a function of whether a prior knowledge explanation was given. When the explanations referenced the cladograms, these students' mean accuracy ($M = 0.70$) was approximately 2.5 times higher than when they referenced prior knowledge ($M = 0.27$), $F(1, 17) = 19.67$, $p < .001$, $MSE = 0.083$, $\eta_p^2 = 0.54$.

Responses to Cladograms That Contradict Prior Knowledge. After identifying the prior knowledge explanations, the coders independently coded those responses from Studies 1 and 2 into Chinn and Brewer's (1998) taxonomy, with each such response being assigned to a single category. The translation from their specific situation to ours is straightforward if "contradictory information depicted in the cladogram" is substituted for "anomalous data." The two coders agreed on the codes for 88.5% of the responses (86% for the college data, 90% for the high school data). Disagreements were resolved by discussion. No responses fell into Chinn and Brewer's categories of professing uncertainty about the validity of the contradictory information, excluding the contradictory information from the domain of the current theory, and holding the contradictory information in abeyance. Also, none were deemed to indicate true change in students' beliefs regarding relationships among the taxa in question. All of the prior knowledge responses fell into Chinn and Brewer's remaining categories, indicating that their taxonomy applies to situations in which the contradictory information is provided diagrammatically and in which students' prior knowledge has the status of a belief rather than a theory.

Table 1 shows the proportion of students who gave at least one prior knowledge response for each of Chinn and Brewer's (1998) remaining four categories and the proportion of all prior knowledge responses placed in each category. At the lowest level, subjects received

TABLE 1
The Proportion of College Students in Study 1 and High School Students in Study 2 Who Provided at Least One Prior Knowledge Response in Each of Four Categories Defined by Chinn and Brewer (1993, 1998), Given That They Gave Any Prior Knowledge Responses at All, and the Proportion of all Prior Knowledge Responses Falling Into Each of These Four Categories

	College Students		High School Students	
Category	Proportion of Subjects	Proportion of Responses	Proportion of Subjects	Proportion of Responses
IGNORE	0.50	0.36	0.88	0.81
REJECT	0.22	0.14	0.12	0.07
REINTERPRET	0.50	0.31	0.12	0.09
PERIPHERAL CHANGE	0.39	0.19	0.08	0.02
Total	18	36	26	81

the *IGNORE* code for their explanation if they ignored the contradictory information (i.e., cladogram structure) and responded based solely on prior knowledge. For example, one student explained her response that *grass* + *geranium* + *mushroom* is the valid biological group for the cladogram in Figure 1b by writing "grows on land & plant." Additional examples from each category are in Table 2. As expected, our students were much more likely than Chinn and Brewer's subjects to receive the *IGNORE* code. Indeed, this was the most frequent type of prior knowledge explanation, accounting for 36% of the responses.

At the next level, students received the *REJECT* code for rejecting the contradictory information as invalid if they explicitly cited problems with the cladogram structure, such as stating that it is wrong or that a taxon in the selected group does not belong. For example, one student explained his choice of the correct biological group for the cladogram in Figure 1b as follows: "Lauren, they share 3 common ancestors. This is obviously false but it is represented in the diagram." Explicit rejections of the contradictory information accounted for 14% of prior knowledge responses.

Chinn and Brewer's (1998) sixth category is to reinterpret the data within one's original theory to retain that theory. We gave the *REINTERPRET* code to explanations that contained elements from both prior knowledge and cladogram structure (i.e., appropriate evolutionary concepts, even if not used appropriately in the student's response) when these elements were simply concatenated rather than integrated. For example, one student identified *seaweed* + *grass* + *geranium* as the valid biological group for the cladogram in Figure 1b because "they're all plants & come from a common ancestor." *REINTERPRET* responses were almost as common as *IGNORE* responses, accounting for 31% of prior knowledge responses.

The remaining responses fit into Chinn and Brewer's (1998) seventh category of peripheral theory change. In this case, students accept the contradictory information as valid and explain the contradiction by making minor changes to their original belief. Two types of responses received the *PERIPHERAL CHANGE* code. First, we included responses that contained elements from both prior knowledge and cladogram structure (like *REINTERPRET*) but that suggested changes to the student's original beliefs because the two sources of information were integrated rather than simply concatenated. For example, for the inference question for the cladogram in Figure 1b, one student wrote that grass and geranium are most likely to share a character with mushrooms because "they are plants that

TABLE 2
Example Prior Knowledge Explanations in Each of Four Categories Defined by Chinn and Brewer (1993, 1998) That Were Observed in Studies 1 (College Students) and 2 (High School Students)

CATEGORY / "Explanation" [Study]	Question: Response Being Explained
IGNORE	
"live in water" [1]; "they all live in water" [2]	Figure 2b clade: beaver, seal, dolphin
"I know they've got teeth" [1]	Figure 3b inference: moose, salmon, bass
"they grow in the ground" [2]	Figure 1b clade: grass, geranium, mushroom
"they are plant like aend [sic] they are like the mushroom" [2]	Figure 1b inference: geranium, grass, seaweed
"they all can be upsid down and carry they're yeung [sic]" [2]	Figure 2b inference: bats, opossums
REJECT	
"Before [for the cladograms with unfamiliar taxa] I would have said Laure[n] (to form the pattern), but that doesn't make any sense. Emily's answer seems to make the most sense. However, I am not sure." [1]	Figure 1b clade: badgers, foxes
REINTERPRET	
"they [moose and lungfish] are most closely related & moose do have enamel although so do sharks." [1]	Figure 3b inference: moose
PERIPHERAL CHANGE	
"While I know frogs/salamanders don't have tooth enamel according to the chart there is a possibility that they do." [1]	Figure 3b inference: frog, salamander, moose

are most closely related to the mushroom." In other words, it is not enough that all the selected taxa are plants (a response that would have received the *IGNORE* code), but from among the plants one must pick the ones that are most closely related to the mushroom (an appropriate evolutionary concept if one accepts that fungi are in the plant group).

The other responses coded as *PERIPHERAL CHANGE* called attention to the conflict between cladogram structure and prior knowledge but ultimately focused on cladogram structure despite misgivings. Consider one student's explanation for her correct inference for the cladogram in Figure 1b that badgers and foxes are most likely to share a character possessed by mushrooms: "According to the digram [sic] since they evolved later. But I didn't think mushrooms & badgers/foxes are really that closely related, it seems more likely the grass, seaweed or geranium would produce chitin." Peripheral theory change accounted for 19% of prior knowledge explanations.

It is interesting to divide the prior knowledge responses into two groups based on whether they provide any evidence that the student also considered evolutionary evidence. *REINTERPRET* and *PERIPHERAL CHANGE* provide such evidence, whereas *IGNORE* and *REJECT* do not. College students' prior knowledge explanations were evenly split between these categories.

STUDY 2

One goal of the present research was to determine the extent to which instructional background or age affects students' responses to diagrammatic information that conflicts with prior knowledge. This is especially important because calls to introduce tree thinking into biology curricula advocate starting in high school or even earlier (e.g., Catley, 2006; Catley, Lehrer, & Reiser, 2005). In Study 2, we attempted to replicate the findings from the matched pairs of cladograms using a sample of high school students. We predicted similar negative effects of depicting relationships among misconceptions taxa on accuracy for the younger students. With respect to the explanations, we expected that (a) appeals to prior knowledge would be more common, (b) lower level categories from Chinn and Brewer's (1998) taxonomy (ignoring or rejecting the contradictory information) would account for a larger proportion of students' responses, and (c) high school students would use fewer different categories of responses.

Method

Subjects. The 35 subjects (20 females, 15 males) were 10th-grade students enrolled in either the basic biology class or the combined chemistry and biology class (a semester of chemistry followed by a semester of biology) at a public comprehensive high school in a small, rural town in a southern Atlantic coastal state. They participated during the middle of fall semester. Their mean age was 15.4 years (range of 15–17). They self-identified their race/ethnicity as White/Caucasian ($n = 32$), Hispanic/Puerto Rican ($n = 2$), and Native American ($n = 1$).

The teacher of these classes reported that students did not have any prior instruction regarding cladograms and tree thinking. In a previous study of tree thinking using a comparable sample of high school students from this same school, Catley, Phillips, and Novick (2013) found that although high school students had greater difficulty answering tree-thinking questions than did college students, it was not an impossible task for them.

Design, Materials, and Procedure. As in Study 1, we manipulated whether the cladograms depicted relationships among unfamiliar or misconceptions taxa within subjects and block order between subjects, with students randomly assigned to the two levels of the between-subjects factor. Seventeen students received the cladograms with misconceptions taxa first and 18 received those with unfamiliar taxa first. The materials consisted of the three matched pairs of cladograms shown in Figures 1–3 plus an additional filler cladogram involving familiar taxa needed to prevent similar cladograms from appearing consecutively (see the Appendix). As in Study 1, two orders of the cladogram pages within each block were used for each block order.

Students spent approximately 35–55 minutes completing two booklets (each for a separate study) and a questionnaire. The questionnaire asked for background information such as sex, age, and year in school. Half of the students ($n = 17$) completed the booklet for this study first. The remaining students completed this booklet after a booklet similar to the one the college students in Study 1 completed first. Students participated in one of several group sessions held in a classroom at their high school but outside the normal school day and were paid $10 for their participation. Students completed the booklets individually and without consulting outside resources.

Results and Discussion

Accuracy. To verify that our tree-thinking task was within the capability of the students in our study, we compared mean accuracy for the multiple-choice clade questions for the cladograms with unfamiliar taxa to what would be expected by chance. Without relevant prior knowledge of the taxa, students either have to reason based on the information shown in the cladograms or guess. As anticipated, the observed accuracy of 0.49 for these questions was significantly higher than what would be expected if students were guessing (0.33), $t(34) = 3.24$, $p < .01$, $SE = 0.05$.

Therefore, as for Study 1, we computed mean accuracy across the clade and inference questions for the matched cladograms with unfamiliar versus misconceptions taxa. These data were analyzed with a 2 (type of taxa; within) × 2 (block order; between) mixed ANOVA.[4] There was a significant main effect of type of taxa, $F(1, 33) = 16.34$, $p < .001$, $MSE = 0.04$, $\eta_p^2 = 0.33$. Students were more accurate for the cladograms with unfamiliar than misconceptions taxa, with means of 0.65 and 0.45, respectively. A follow-up between-subjects ANOVA on just the first block of cladograms showed the same pattern of results: $F(1, 33) = 10.89$, $p < .01$, $MSE = 0.06$, $\eta_p^2 = 0.25$, with means of 0.67 and 0.39, respectively. These results replicate Study 1, although the accuracy scores are much lower. High school students' lower accuracy reflects their greater use of prior knowledge, as discussed in the next section.

Unlike in Study 1, there was no effect of block order, $F(1, 33) = 1.10$, $p > .30$, $MSE = 0.09$, $\eta_p^2 = 0.03$. The interaction also was not significant, $F(1, 33) = 0.55$, $p > .45$, $\eta_p^2 = 0.02$. The block order effect in Study 1, without an interaction, indicates that college students were able to apply a strategy adopted initially to answer the same kinds of questions presented subsequently. This carry-over effect was either positive or negative depending on which cladograms students received first. High school students, however, were unable to carry forward their more successful strategy for answering questions about cladograms with unfamiliar taxa to improve their reasoning for cladograms with misconceptions taxa. Whenever prior knowledge about the taxa was available, it was heavily weighted in their reasoning. This difference between the two groups may be related to 10th graders' lower accuracy for the cladograms with unfamiliar taxa: $M = 0.65$ versus 0.84. Successful application of a prior strategy to a new situation is more likely with material students understand fairly well than with material for which their understanding is less robust.

Prior Knowledge Explanations

Frequency of Occurrence. Prior knowledge explanations were given by 26 students. This is a much higher percentage of students (74%) than was found in the college sample (26%), $\chi^2(1, N = 105) = 22.61$, $p < .001$, $\phi = .46$. These 26 students gave an average of 3.12 prior knowledge explanations each, which is significantly more than the mean of 2.06 for the college students, $F(1, 42) = 4.75$, $p < .05$, $MSE = 2.51$, $\eta_p^2 = 0.10$. Thus, as predicted, 10th graders were more likely to admit responding based on prior knowledge than were college students. Replicating the college student results, the prior knowledge responses occurred for cladograms with misconceptions taxa almost exclusively. Such responses comprised 39% of 10th-graders' explanations for those cladograms but only 2% of their explanations for cladograms with unfamiliar taxa. All of the latter explanations were for the spiders cladogram (Figure 1a), which had common names. For example, two students chose *crab spider + jumping spider + orchard spider* as the valid biological

[4]Preliminary analyses indicated no effect of whether the booklet for this study was completed first or second.

group "because they all have to do with an outdoor spider" or "because spider [sic] do or live in these." There were no such explanations for questions concerning either yeasts or streptococcus bacteria, which had Latin names.

The statistical analysis comparing accuracy for questions for which prior knowledge explanations were or were not given was based on 23 of the 26 students who gave such responses because three students referenced prior knowledge for all six questions about the cladograms with misconceptions taxa. Replicating the pattern found for college students, when high school students' explanations did not refer to prior knowledge, their mean accuracy score ($M = 0.65$) was approximately 4.5 times higher than when they did refer to prior knowledge ($M = 0.14$), $F(1, 22) = 34.11$, $p < .001$, $MSE = 0.09$, $\eta_p^2 = 0.61$.

Responses to Cladograms With Misconceptions Taxa. High school students' prior knowledge explanations for cladograms with misconceptions taxa ($N = 81$) were found to fall into the same four of Chinn and Brewer's (1998) categories as college students' responses: ignore the contradictory information, reject the contradictory information, reinterpret the contradictory information while retaining the original belief, and peripheral belief change. The proportions of (a) students who gave at least one explanation in each category and (b) all prior knowledge responses that fit into each category are shown in Table 1. What is most striking is high school students' extremely strong preference to ignore the contradictory information. Such explanations were given by 88% of students who referred to prior knowledge, and 81% of all prior knowledge responses involved ignoring the contradictory information. For example, for the cladogram in Figure 3b, four students said shark and moose were most likely to have tooth enamel like the lungfish because "they have teeth." Another five students used the presence of teeth to justify inferences to a variety of other taxa. Additional examples in this category are given in Table 2.

We made two predictions with respect to the comparison between college and high school students' responses to misconceptions taxa. First, Chinn and Brewer (1998) suggested that younger students might produce a narrower range of responses than undergraduates. We tested this hypothesis by giving each student who produced at least one prior knowledge response a score reflecting the number of different categories into which those responses fell. For example, a student who gave one or more responses coded *IGNORE* and one or more responses coded *REINTERPRET* received a score of 2. A student whose prior knowledge responses were all coded *IGNORE* received a score of 1. This scoring procedure biases against finding the predicted difference because college students gave fewer prior knowledge responses per person, which makes it difficult for them to show the predicted greater diversity of types of responses. Nevertheless, the statistical analysis indicated that, on average, college students gave more different kinds of explanations ($M = 1.61$) than did 10th graders ($M = 1.19$), $F(1, 42) = 4.80$, $p < .05$, $MSE = 0.39$, $\eta_p^2 = 0.10$.

We further predicted that high school students would give relatively more responses at lower levels in Chinn and Brewer's (1998) taxonomy, whereas college students would give more responses at higher levels. Because students gave differing numbers of prior knowledge responses of differing types for different questions, the simplest method for testing this hypothesis was to assign students a score reflecting the overall quality of their prior knowledge explanations. To accomplish this, we assigned *IGNORE, REJECT, REINTERPRET,* and *PERIPHERAL CHANGE* responses, respectively, scores of 0–3 and then computed an average explanation quality score for each student. Although the quality of prior knowledge explanations was low overall, as the data in Table 1 clearly show, college students had much higher quality scores ($M = 1.45$) than did high school students ($M = 0.38$), $F(1, 42) = 18.53$, $p < .001$, $MSE = 0.66$, $\eta_p^2 = 0.31$.

In sum, 10th-graders' explanations for their answers to tree-thinking questions were less sophisticated than those of college students in two respects. First, they were much more likely to provide explanations that called upon prior knowledge, rather than responding based on the information provided in the cladograms as requested. Second, comparing the prior knowledge responses given by the two samples, those of the high school students were both less varied and less sophisticated. As a general rule, they simply ignored the contradictory information provided by the cladogram and responded based solely on prior knowledge.

STUDIES 3A AND 3B

In Studies 1 and 2, we kept the cladogram structure constant and varied the taxa to which that structure was applied. In Study 3, we adopted the complementary strategy of varying the cladogram structure relevant to a particular misconception—that birds are not reptiles. In an unpublished study, the present authors asked a group of 71 college students recruited from the same source as the students in Study 3a to complete a tree-thinking assessment. Question 15 presented two three-taxon cladograms showing possible evolutionary relationships among mammals, birds, and snakes. In one cladogram, birds were shown as more closely related to mammals; in the other, they were shown as more closely related to snakes. Students were asked which cladogram shows the correct evolutionary relationships among these taxa. Only 52% picked the scientifically accepted cladogram. Either approximately half of our college student population knows that birds are more closely related to snakes than to mammals, and the other half "knows" that birds are more closely related to mammals than to snakes, or these students have no idea which set of relationships is correct and so guessed.

Even under the first, more optimistic interpretation, though, knowing that birds are more closely related to snakes is not the same as placing birds in the same taxonomic category as snakes. For example, although students presumably (correctly) know that rodents are more closely related to felines than to birds, the correctness of such a relationship does not mean that rodents are carnivores. As noted earlier, people think that the most inclusive meaningful groups of animals are those at the folk-biological rank of life form, such as land mammals, birds, "reptiles,"[5] amphibians, fish, and insects (e.g., Atran, 1998; Berlin et al., 1973). In folk-biological classification, birds and "reptiles" are seen as distinct groups of animals: A robin can no more be a reptile than can a squirrel. Supporting this misconception from folk-biological taxonomy, middle school life science texts typically discuss birds and "reptiles" in separate chapters, thereby strongly (and incorrectly) implying that birds are not reptiles (e.g., Life Science, 2007). Moreover, the second author, who interacts with and teaches biology undergraduates on a daily basis, still encounters students who find the concept of birds as reptiles troubling even when presented with unique shared character evidence such as scales, feathers, and wish bones. Clearly, birds' supposed nonreptilian status is a widely and deeply held misconception.

In Study 3a, we presented relatively weak structural evidence supporting the contradictory, but scientifically accepted, classification that birds in fact are reptiles. Given the weakness of the evidence, we predicted that most students would decline to classify birds as reptiles. As Chinn and Brewer (1993) noted, scientists do not reject a current theory in favor of a new one unless there is strong evidence contradicting the old theory and supporting the new theory. Accordingly, Study 3b presented stronger evidence supporting

[5] We placed *reptiles* in quotes because the folk-biological category is not a meaningful biological group (i.e., is not a clade) as it excludes birds.

the classification of birds as reptiles. If college students are sensitive to the strength of evidence supporting a conclusion that contradicts their prior knowledge, those in Study 3b should be more likely than those in Study 3a to classify birds as reptiles. In each study, we examined students' willingness to endorse this classification, as well as their reasons for either accepting or rejecting it.

Method

Study 3a. Seventy college students received the cladogram shown in Figure 4a as part of the booklet containing the matched pairs of cladograms reported in Study 1. This cladogram shows that birds are more closely related to snakes and lizards than are turtles because birds share a more recent common ancestor with snakes and lizards than do turtles. Like the cladograms in the two previous studies, this cladogram was prefaced by a statement calling attention to its provenance in evolutionary biology to "pass the test of credibility" (Chinn and Brewer, 1993, p. 24) and cue students that we were requesting answers based on current scientific classification. The first question was a slightly modified version of that used for the matched cladograms: "The following students disagree about

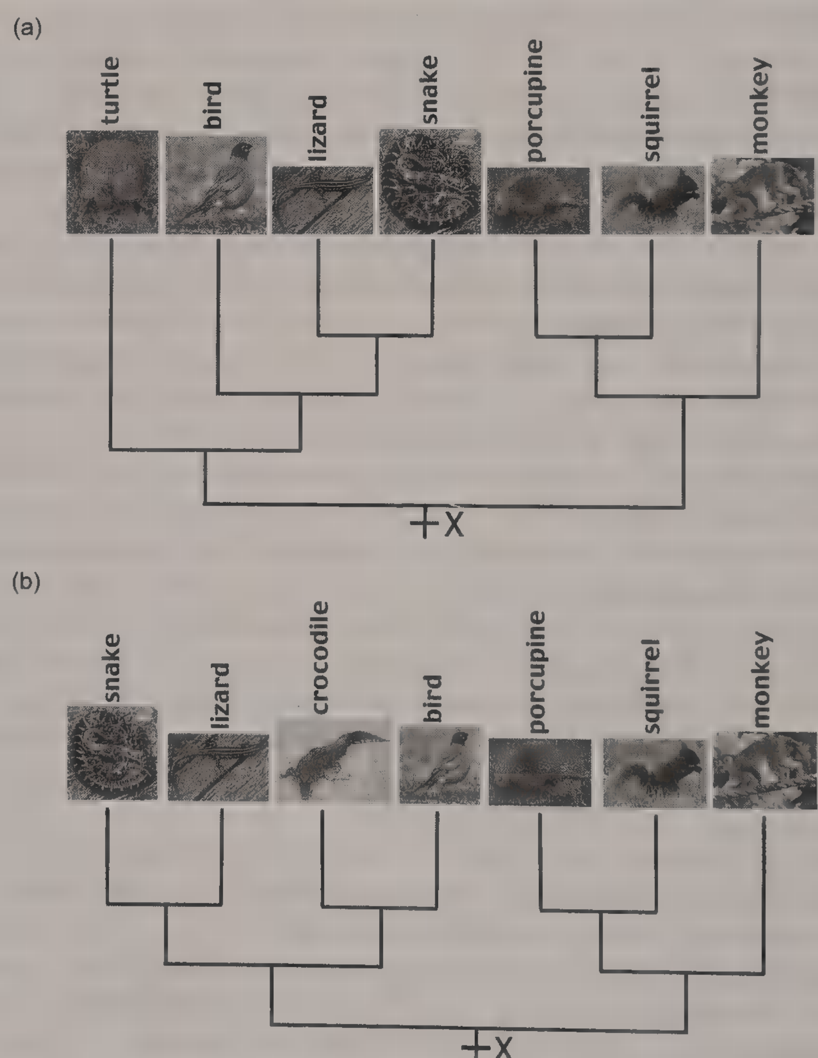


Figure 4. Cladograms presenting evidence supporting the conclusion that birds are reptiles. The pictures were presented in color in the materials subjects received. (a) Weaker evidence version used in Study 3a. (b) Stronger evidence version used in Study 3b.

which taxa should be considered reptiles. Which student's definition of reptiles best reflects evolutionary evidence?" The three named students said, respectively, that reptiles are (a) lizards and snakes; (b) turtles, lizards, and snakes; and (c) turtles, birds, lizards, and snakes. The second option is consistent with what students likely learned in high school (e.g., Miller & Levine, 2002). From a cladistic perspective, the first and third response options are equally defensible answers because both comprise a clade. Because turtles traditionally are considered reptiles, the third response, which includes birds, is arguably the best answer (e.g., Freeman, 2011; Reece et al., 2011). Students were asked to explain their definition choice.

The set of relationships shown in Figure 4a provides relatively weak evidence for classifying birds as reptiles (for nonbiologists) for two reasons. First, turtles are not seen as good members of the reptile category in folk-biological classification. Shapiro and Palermo (1970) asked college students to list the first four items they thought of as representative reptiles. Snake and lizard were listed most often, being generated by 81% and 63% of subjects, respectively. Turtle, in contrast, was listed by only 26% of subjects. Thus, turtles are viewed as atypical reptiles (Mervis, Catlin, & Rosch, 1976). Trowbridge and Mintzes (1988) directly asked college students in a nonmajors introductory biology course whether lizards, snakes, and turtles are reptiles. These taxa were classified as reptiles by 78%, 77%, and 33% of students, respectively. Turtles were instead classified as amphibians by 52% of students. Because those students might have been thinking of sea turtles, we used a picture of a land turtle in our cladogram. Second, structurally, birds are not in the same immediate group as snakes and lizards, which are both typical reptiles. Given this weak evidence, coupled with a strong prior belief that birds are not reptiles, we predicted that students would prefer "reptiles = snakes + lizards" to the definition that also included turtles and birds. Students' explanations provided direct evidence concerning whether their choice of snakes and lizards only was due to the misconception that birds are not reptiles.

Study 3b

Materials. The students in this study received the cladogram shown in Figure 4b (again prefaced by a statement indicating its provenance in evolutionary biology), which included crocodiles rather than turtles. Crocodilians (crocodiles, alligators, caimans, gharials) and birds are more closely related to each other than either group is to any other extant group of animals. In Shapiro and Palermo's (1970) study, students more often listed crocodiles (38%) than turtles (26%) when asked to list representative reptiles. Moreover, *The Crocodile Hunter* was a very popular TV show that aired from 1997 to 2004, when the students in our study would have been in late elementary through early high school. Greater frequency of occurrence increases people's judgments of how typical an item is of its category (Nosofsky, 1988; Novick, 2003). Students received the same reptile definition choices as in Study 3a, with *crocodiles* substituted for *turtles*: (a) *reptiles = snakes + lizards*, (b) *reptiles = snakes + lizards + crocodiles*, and (c) *reptiles = snakes + lizards + crocodiles + birds*. As in Study 3a, students had to explain their answer.

The set of relationships shown in Figure 4b provides stronger evidence for classifying birds as reptiles: (a) Structurally, birds are immediately linked to crocodiles, which argues strongly for treating these two taxa similarly, and (b) semantically, nonbiologists think crocodiles are better examples of reptiles than are turtles. Thus, we expected students to be less willing to exclude crocodiles than turtles from the definition of reptiles and, therefore, more willing to include birds.

Subjects, Design, and Procedure. The subjects were 124 undergraduates (67 females, 54 males, 3 undisclosed sex) from the same institution, and recruited from the same source, as the students in Study 3a. They were participating in a study testing the effectiveness of a new tree-thinking instructional booklet (Novick, Schreiber, & Catley, 2014) and were paid $25 for their participation. Their average year in school was 2.70 (2 = sophomore, 3 = junior). We added the birds are reptiles cladogram to the test booklet for that study.

For the instructional study, students were divided into two groups based on their responses to the biology coursework question (identical to that used in Study 1). Those who had taken both semesters of the two-semester introductory biology sequence for biology majors and pre-med students were classified as having a stronger biology background ($n = 63$). They had received a day or two of instruction related to cladograms and tree thinking in the second semester class. (Students who had taken the evolution class were excluded from the study.) All others were classified as having a weaker biology background ($n = 61$).[6] Their background in tree thinking is comparable to that of the students in Study 1 (and 3a). The stronger background students had taken a mean of 2.14 (range of 2–4) of the semester-long biology classes listed on our questionnaire, compared with a mean of 0.32 (range of 0–1.5) for the weaker background students.

Within each biology background group, approximately half of the students were randomly assigned to complete a self-paced instructional booklet that taught core aspects of tree thinking, such as the concepts of most recent common ancestry and clades and the basis for determining relative evolutionary relatedness.[7] There was no information about the classification of birds and, therefore, no information that would help students choose between the two valid responses for the reptile definition question, both of which comprise a clade. The instruction would be expected to help students choose one of these two responses over the invalid response, but as almost nobody in Study 3a chose the invalid response, as we discuss in the Results section, we expected students in the instructional and control conditions to perform similarly for the reptile definition question.[8] Students in the instructional condition spent approximately 30 minutes reading this booklet and taking two practice quizzes on the content of the booklet. Students in the control condition spent the same amount of time taking several individual differences tests. Then all students completed two problem booklets. The reptile cladogram was on page 8 of the 18-page first problem booklet. Finally, all students completed an individual differences test and the background questionnaire. Students participated alone or in the presence of one or more other students in a single session that lasted approximately 2 hours. They completed the tasks individually and without consulting outside materials (including the instructional booklet in that condition).

Preliminary analyses of the data from Study 3b indicated that students' responses to the reptile definition question did not vary as a function of either biology background or condition.[9] Given these results, we treated the sample as a single group in our analyses.

[6] We had to work very hard to recruit stronger background students into the paid subject pool to get such students in this study. Without active recruitment, there are few such students in the pool. That is why the mean number of biology classes for the Study 1 sample was only 0.59 and why we did not split that sample into two groups based on biology background.

[7] The current version of the instructional booklet can be downloaded from the first author's Web site: http://www.vanderbilt.edu/peabody/novick/novick_home.html.

[8] Obviously, we expected (and, indeed, found) differences between the two conditions for the test problems that actually tested concepts that were targeted in the instruction (Novick et al., 2014).

[9] The instruction result was predicted, as we already discussed. The absence of an effect of biology background is more surprising, as we might have expected the stronger background students to have

Results and Discussion

Defining Reptiles. In both studies, very few students chose the reptile definition that was not a valid biological group. In Study 3a, five of the 70 students (7%) responded that reptiles are snakes, lizards, and turtles. In Study 3b, three of the 124 students (2%) chose the comparable snakes, lizards, and crocodiles definition. We turn, then, to considering students' preferences between the two valid groups (i.e., the two correct answers). Two students in each study wrote that either or both of these responses is correct. Because these students failed to state a preference, we focus our analyses on the remaining students ($n = 63$ in Study 3a, $n = 119$ in Study 3b) who chose either the definition that excluded turtles/crocodiles and birds or the definition that included those taxa.

In each study, students had to weigh the evidence for the scientifically accepted conclusion that birds are reptiles against their prior knowledge that birds are birds and therefore cannot be reptiles, which constitute a different life form. If students realize the contingency between including or excluding turtles/crocodiles and birds, the least belief-damaging reconciliation of the conflicting information, what Chinn and Brewer (1993, 1998) referred to as reinterpreting the data, is to remove turtles/crocodiles from the reptile category because "everyone knows" birds are not reptiles. Consider the following response from Study 3a: "Now I am using common sense → lizards and snakes are reptiles while birds definitely are not. I would think turtles would be because of their appearance, but apparently birds evolved from the same ancestor as turtles." (For students who think box turtles are amphibians, no reconciliation would be needed. Only two of the 70 students in Study 3a wrote that turtles are amphibians, although of course others might have held the same belief.)

Because changing one's core beliefs is difficult, and the evidence supporting the classification of birds as reptiles was weak in Study 3a, we predicted that students in that study would be more likely to resolve the discrepancy between prior knowledge and the evolutionary information provided in the cladogram by excluding rather than including both turtles and birds from the definition of reptiles. A binomial test indicated that turtles and birds were much more likely to be excluded ($n = 50$, 79%) than included ($n = 13$), $p < .001$, as predicted.

If college students are sensitive to the strength of the evolutionary evidence depicted in cladograms, those in Study 3b should have found including birds a more persuasive option than did those in Study 3a. Two analyses supported this hypothesis. First, a binomial test on the data from Study 3b indicated no significant difference in the frequency with which the two definitions were chosen, $p > .55$, with 53% of students selecting the exclusion (snakes and lizards) definition. Second, a chi-square test comparing the distributions of definition choices across the two studies was significant, $\chi^2(1, N = 182) = 12.22, p < .001$, $\phi = .91$. Although the stronger evidence depicted in the cladogram in Figure 4b compared with that in Figure 4a reduced the percentage of exclusion definitions, the fact that half of the students in Study 3b still endorsed that definition makes it clear that birds' supposed nonreptilian status is a very strongly held belief.

Prior Knowledge Explanations. To better understand the bases for endorsing the different definitions, the explanations were coded into four categories. The *NOT_REPTILE*

encountered the idea that birds are reptiles in their introductory biology class for majors. If they were exposed to this contemporary and well-supported classification, they evidently either forgot it or rejected it.

TABLE 3
The Relation Between Reptile Definitions and Prior Knowledge Explanations for the Subset of College Students in Studies 3a and 3b Who Chose One of the Two Valid Biological Groups and Whose Explanation for That Choice Referred to Prior Knowledge

	Explanation Code		
Definition	NOT_REPTILE	MUST_REPTILE	PRIOR_KNOW
Study 3a			
Taylor (snakes, lizards)	15 (75%)	1 (5%)	4 (20%)
Jordan (snakes, lizards, birds, turtles)	1 (20%)	4 (80%)	0 (0%)
Study 3b			
Taylor (snakes, lizards)	27 (87%)	3 (10%)	1 (3%)
Jordan (snakes, lizards, birds, crocodiles)	1 (5%)	16 (76%)	4 (19%)

Notes: The exclusion definition was provided by Taylor, and the inclusion definition was provided by Jordan.
Cell entries are frequencies (percentages of row totals)

code was given to students who wrote that birds are not reptiles. This "fact" had to be directly stated or *very* strongly implied, based on reasoning from prior knowledge. In Chinn and Brewer's (1993, 1998) taxonomy, these responses could involve either rejecting or reinterpreting the contradictory information depicted in the cladogram. Because the rejection interpretation is most plausible if those explanations supported the definition that reptiles are snakes, lizards, and turtles/crocodiles, and almost nobody selected that definition, we believe these responses indicate reinterpretation.

Students received the *MUST_REPTILE* code if they wrote that (a) birds must be reptiles if turtles/crocodiles are, (b) birds must be reptiles because they share the same ancestor as reptiles, or (c) birds and turtles/crocodiles must both be included (excluded) if one of them is. It was not sufficient just to say that snakes, lizards, turtles/crocodiles, and birds are all reptiles. Students had to convey the idea of contingency, that including/excluding one taxon requires the same response for another taxon. These responses qualify as true "theory" change in Chinn and Brewer's (1993, 1998) scheme, although, as we will see, a few students expressed some skepticism about this response. *NOT_REPTILE* took precedence over *MUST_REPTILE* if both codes applied.

Responses that did not meet the criteria for either of these two codes but that nevertheless referred to prior knowledge received the *PRIOR_KNOW* code. All other responses were coded *OTHER*. The same two coders, working independently, agreed on the coding of 91% of the explanations from each study. Disagreements were resolved by discussion.

Despite being told to respond based on the evolutionary evidence provided in the cladogram, 40% of students in Study 3a and 44% of those in Study 3b who selected either the exclusion or inclusion definition of reptiles received one of the three prior knowledge codes. In each study, the distribution of definition choices by students who gave prior knowledge explanations mirrors that seen in the full sample. Table 3 shows the relation between

students' definition choices and their (subsequent) explanations based on prior knowledge for both studies.

Although students' definition choices varied across the two studies, their explanations for these choices were quite similar. In separate chi-square tests, we compared the prior knowledge explanations given for the exclusion and inclusion definitions across the two studies. Both tests failed to show a significant difference between the two studies: $\chi^2(2, N = 51) = 4.04$, $p > .10$, Cramer's $\phi = .28$, and $\chi^2(2, N = 26) = 2.18$, $p > .30$, Cramer's $\phi = .29$, respectively. Both studies, however, found a significant relationship between the definition chosen and the explanation given to support that definition: $\chi^2(2, N = 25) = 14.14$, $p < .001$, Cramer's $\phi = .75$, for Study 3a; and $\chi^2(2, N = 52) = 34.18$, $p < .001$, Cramer's $\phi = .81$, for Study 3b.

In both studies, students who defined reptiles as snakes and lizards, and who referred to prior knowledge to explain this choice, overwhelmingly wrote that birds are not reptiles, thus preserving their original belief about birds and reinterpreting the status of turtles/crocodiles: $M = 81\%$. Examples from Study 3a include (a) "birds are not considered reptiles" and (b) "Robin's is a bad answer because she skipped birds illogically. Jordan's answer could be true it is a valid grouping, but Taylor's seems the best because it is valid & does not include animals not normally considered reptiles, like birds." Examples from Study 3b include (a) "Crocodiles cannot be included without birds, and since birds are not reptiles, crocodiles cannot be in the reptile group, only snakes & lizards" and (b) "Snakes & lizards are most closely related → reptiles definitely do not include birds." The five students who supported the exclusion definition by other appeals to prior knowledge wrote that snakes and lizards are reptiles or have reptilian characteristics.

In contrast, students who defined reptiles as including all four appropriate taxa overwhelming explained this definition by appealing to the contingency between turtles/crocodiles and birds: $M = 78\%$. These explanations suggest true belief change. Examples from Study 3a include (a) "Since turtles lizards & snakes are most commonly thought of as reptiles, birds must also be included because they share a common ancestor w/ lizards & snakes" and (b) "according to the diagram birds are more similar to lizards and snakes than turtles. Therefore if a turtle is a reptile, so is a bird." Examples from Study 3b include (a) "Since snakes, lizards, and crocodiles are all clearly reptiles, their clade must all be reptiles, given their MRCA so birds must be included" and (b) "I already knew snakes, lizards, and crocodiles were reptiles so if birds are next to crocs then they are reptile too." A few students in Study 3b, however, hedged a bit in their *MUST_REPTILE* explanation: (a) "A crocodile, snake + lizard are reptiles. However the crocodile is more closely related to the bird than either of the other two—birds are also reptiles (by the diagram)" and (b) "These all stem from the same original branch. Because the first three are considered reptiles by today's standard, evolution states that birds must have also descended from the same, according to this cladog." These students seem to leave open the possibility that a different cladogram might give a different (more familiar) answer.

We should also note that three of the four *PRIOR_KNOW* explanations supporting the inclusion definition in Study 3b potentially could be considered *MUST_REPTILE* explanations. These students wrote that one needs to make a clade and that crocodiles must be included, which implies birds must also be included, e.g., "only of 3 groups that designates a clade that also includes crocodiles (given that you know these are reptiles)." We coded *MUST_REPTILE* conservatively, requiring students to explicitly mention the conclusion that birds are reptiles.

GENERAL DISCUSSION

Responses to Contradictory Information From Prior Knowledge Versus Depicted in Tree-of-Life Diagrams

In many areas of science, critical information is presented diagrammatically. Yet, little is known about how students respond to such information when it conflicts with their prior knowledge. We examined students' responses to contradictory information presented in tree-of-life diagrams (cladograms), which are standard currency in contemporary evolutionary biology.

Effects of Prior Knowledge While Keeping Branching Structure Constant. In Studies 1 and 2, college students and 10th graders, respectively, answered questions about three matched pairs of cladograms (see Figures 1–3). Both cladograms in a pair had the same nested structure but differed in whether they depicted relationships among unfamiliar taxa or familiar taxa about which students are known to have misconceptions concerning their relationships. As predicted, students had lower accuracy for the cladograms with misconceptions taxa even though the cladograms and questions (about most recent common ancestry and valid biological groups) for each pair were structurally identical. This was a medium-size effect for college students and a large effect for 10th graders.

Reduced accuracy for the cladograms with misconceptions taxa is one indicator that many students weighted their prior knowledge more heavily than cladogram structure (i.e., evolutionary evidence). In addition, both groups of students gave prior knowledge explanations (almost) exclusively for the cladograms with misconceptions taxa, and these explanations were much more likely to accompany incorrect than correct answers. These results suggest that textbook authors and instructors need to consider carefully which taxa to include in cladograms that are used to teach core tree-thinking skills. We discuss this issue in the section on implications for instruction.

Despite a large body of research in education and psychology that has documented the benefits of visual over verbal representations for learning, reasoning, and problem solving (e.g., Ainsworth & Loizou, 2003; Hegarty & Just, 1993; Kindfield, 1993/1994; Rotbain et al., 2006; Sweller et al., 1990), our results indicate that diagrammatic depictions of information that conflict with students' common misconceptions are not powerful enough, by themselves, to overcome students misconceptions. Nevertheless, it remains possible that diagrammatic depictions are more successful in this regard than are verbal ones. We did not conduct such a comparison because verbal depictions of nested evolutionary relationships are very difficult to understand, so it did not seem a fair test. It would be useful to address this issue in future research using different science content for which easy-to-understand verbal and diagrammatic representations can be created.

Differences Between High School and College Students. Five aspects of our results suggest that the existence of misconceptions had a larger negative effect on high school than college students' ability to engage in tree thinking. First, the effect size for the manipulation of misconceptions versus unfamiliar taxa was larger for the 10th graders. Second, 10th graders were more likely to explicitly refer to prior knowledge in explaining their responses to the tree-thinking questions. Third, among students who gave prior knowledge explanations, 10th graders gave more such explanations per person than did college students. Fourth, based on a coding of the prior knowledge responses into Chinn and Brewer's (1993, 1998) categories, 10th graders gave fewer different types of responses per person than did college students. Finally, 10th graders' responses were predominantly

(88%) classified at lower levels in Chinn and Brewer's taxonomy (*IGNORE* or *REJECT* the contradictory information), whereas college students' explanations were evenly split between those suggesting that they also considered evolutionary evidence (*REINTERPRET* the contradictory information, *PERIPHERAL BELIEF CHANGE*) and those suggesting that they did not (*IGNORE, REJECT*).

The conclusion about differences between high school and college students must be considered somewhat tentative, however, because of the different populations from which the two groups were sampled. The college students were recruited from a highly selective private university. The high school students, in contrast, were recruited from a rural school from which students are more likely to attend a community college than a 4-year college or university, and only a tiny percentage attend a private university. It is possible that the differences we observed reflect the educational opportunities of the students rather than year in school. In that case, college students sampled from a community college or regional state university might respond more like our high school sample, and high school students sampled from a top academic high school might respond more like our college students.

Regardless, it is important to keep in mind that the college students in our studies, despite their strong academic backgrounds, showed similar negative effects of being asked to reason about cladograms depicting scientifically accepted relationships among taxa that contradict common misconceptions about those relationships. Moreover, despite high school students' lower overall accuracy at answering the tree-thinking questions compared with college students, they did show some ability to appropriately use evolutionary evidence depicted in cladograms. Like the college students, they had significantly higher accuracy scores for the cladograms with unfamiliar than misconceptions taxa.

Overall, our results suggest that tree-thinking instruction for both groups of students, but especially for high school students, needs to include information about the nature of the evidence supporting the nested structure depicted. For example, the difference between (derived) shared characters that are the result of most recent common ancestry, and thus inform evolutionary relationships, and those that are the result of independent evolution from different ancestors must be stressed. Many student misconceptions about relationships among taxa are due to reliance on similarities that result from convergent evolution. For example, beavers, seals, and dolphins are often judged to belong in the same group because they share an aquatic habitat; reptiles are believed to include only cold-blooded (ectothermic) animals. Although such similarities may be informative about present-day ecological relationships, they are uninformative with respect to historical evolutionary relationships.

Effects of Branching Structure on Reasoning About a Particular Misconception. In Studies 1 and 2, we kept the cladogram structure constant and varied the taxa to which that structure was applied. In Studies 3a and 3b, we adopted the complementary strategy of varying the cladogram structure relevant to a particular misconception—that birds are not reptiles. We gave college students evolutionary evidence that birds in fact are reptiles and asked them to choose which of three definitions of reptiles is consistent with this evidence. Students clearly attended to the scientific evidence: In both studies, few students chose the group that is consistent with common knowledge about taxa that are reptiles (*snakes + lizards + turtles/crocodiles*) but inconsistent with the evolutionary evidence shown in the cladogram. Rather, almost every student chose one or the other of the two valid biological groups: snakes and lizards or those taxa plus turtles/crocodiles and birds. When faced with the weaker evidence concerning the reptilian status of birds in Study 3a (see Figure 4a), these latter students overwhelmingly (79%) preferred the definition that excluded birds (i.e.,

consisted of just snakes and lizards) and therefore was consistent with common knowledge that birds are not reptiles. Given the stronger evidence in Study 3b (see Figure 4b), however, only about half (53%) of these students chose the more restrictive definition.

Although it is encouraging that college students were sensitive to the strength of the evidence contradicting the misconception that birds are not reptiles, two results temper this sentiment. First, half of the students who faced the dilemma that preserving the belief that birds are not reptiles meant giving up the belief that crocodiles are reptiles chose to exclude rather than include both of those taxa. Second, although half of the students who received the stronger evidence had taken the year-long introductory biology class for majors, their responses were indistinguishable from those who had not taken that class. The "fact" that birds are not reptiles appears to be entrenched and resistant to change. Perhaps if birds were introduced as dinosaurs when young children (at least those in the United States) are fascinated by those taxa, it would be easier for students to later think of birds as reptiles.

Students' explanations clarified the roles of prior knowledge and evolutionary evidence in supporting these definitions. Across both studies, 82% of students who gave a prior knowledge explanation of their conclusion that reptiles are snakes and lizards only justified this choice by stating that birds are not reptiles. These students appear to have reinterpreted their definition of reptiles to exclude turtles/crocodiles so as to be able to exclude birds also. We believe these responses involve reinterpretation rather than rejection of the contradictory information depicted in the cladogram because if turtles/crocodiles were not originally believed to be reptiles, there would be no reason to explain that birds are not reptiles; one could simply say that snakes and lizards are the only reptiles for any of a variety of prior knowledge reasons (e.g., they both are cold-blooded, they both have scales). In contrast, for students who gave a prior knowledge explanation to support the definition of reptiles as snakes, lizards, turtles/crocodiles, and birds, 77% justified this definition by stating that if turtles/crocodiles are reptiles then birds must be too. These students appear to have changed their prior belief concerning the taxonomic status of birds given the evolutionary evidence provided.

Biologists, using the latter type of reasoning, find the evidence that birds are reptiles more persuasive than do students (e.g., Freeman, 2011; Lee et al., 2004; Reece et al., 2011; Thanukos, 2009). Of course, this may be because biologists are aware of even stronger evidence for classifying birds as reptiles. Although we presented comparative evidence involving turtles in Study 3a and crocodiles in Study 3b, evolutionary biologists know about both sets of evidence and more. Presenting students with a combined cladogram for which preserving their prior belief that birds are not reptiles would require excluding turtles, crocodiles, and possibly other known reptiles (e.g., alligators) from the group might be an informative way both to (a) test how entrenched versus malleable is their belief about birds' nonreptilian status and (b) convince them to change their conception of birds. Taking a cue from the response of one student in Study 3b, nonavian dinosaurs (e.g., *Tyrannosaurus rex, Stegosaurus*) also could be added to the cladogram to bolster birds' reptilian status. It would also be interesting to examine high school students' sensitivity to the strength of evidence supporting a conclusion that contradicts a strongly held misconception.

Further research on this topic would profit from adopting a conceptual change framework in which students are evaluated in advance concerning their knowledge of the reptilian status of birds. The effectiveness of a variety of methods for teaching students the scientifically accepted classification could then be compared for students who strongly believe that birds are not reptiles versus are unsure about the classification of birds. For example, the latter group might know that birds are dinosaurs but be unsure whether this means they therefore must be reptiles.

Implications for Teaching Tree Thinking

Our research was motivated in part by recent calls to include tree thinking (i.e., cladograms and how to interpret and use them) in biology curricula at both the college and high school levels (e.g., Baum et al., 2005; Catley, 2006; Catley et al., 2005; Gilbert, 2003; Goldsmith, 2003; O'Hara, 1988). The results of our studies suggest three implications for designing initial instruction in tree thinking.

First, it may be preferable to use cladograms with unfamiliar taxa when introducing the principles of tree theory because such cladograms will not cue seemingly relevant folk-biological information in memory and therefore should have no possibility of presenting relationships that contradict students' prior knowledge. Both high school and college students responded more accurately to our tree-thinking questions for the cladograms with unfamiliar taxa. Note that the unfamiliar taxa need not be unpronounceable Latin names as in Figures 2a and 3a. They could be common names of taxa about which students have no information about their relationships, such as kinds of spiders (Figure 1a), insects, birds, or plants. Although the college students in Study 1 were able to carry forward appropriate cladogram interpretations based on the nested branching structure that they adopted for the cladograms with unfamiliar taxa to the more challenging situation in which the cladograms presented relationships among misconceptions taxa (a medium-size effect), high school students were not. It is important to note, however, that we did not provide any instruction in Studies 1 and 2. We presume that application of an appropriate strategy to new, and more demanding, content can be obtained for both groups with appropriate instruction. Indeed, in both studies, within-subjects comparisons showed that students were much more successful at answering tree-thinking questions about cladograms with misconceptions taxa when their explanations referenced the cladogram rather than prior knowledge.

Second, if familiar taxa are used in initial instructional examples, one should endeavor to ensure that the relationships depicted do not contradict students' prior (incorrect) knowledge so that students can attend to the principles and underlying theory being taught without distraction from misconceptions that may lead them to discount the instruction. This strategy will require additional knowledge concerning students' misconceptions. Although some data on this topic exist (e.g., Johnson, Mervis, & Boster, 1992; Morabito et al., 2010; Novick, Catley, & Funk, 2011; Trowbridge & Mintzes, 1988), not all of the studies address this question from an evolutionary perspective.

Third, students need to be taught why cladograms provide a strong source of evidence concerning historical evolutionary relationships that should be weighted more heavily than uninformative taxon similarities (e.g., of habitat, mode of locomotion, or thermoregulatory mechanisms, to cite some of the explanations given in our prior research) that may often reflect convergent evolution rather than shared ancestry (also see Morabito et al., 2010). Knowledge acquisition in biology (in part) means understanding the evolutionary basis for grouping taxa and being able to distinguish taxon similarities that reflect the structure of the domain (i.e., are due to shared ancestry and thus are informative) from those that are superficial (i.e., due to convergent evolution from separate ancestors and thus uninformative with respect to evolutionary relationships). Learning the evidential basis underlying cladograms should occur as part of a comprehensive curriculum that integrates tree thinking with nature of science concepts more generally (Catley, Novick, & Funk, 2012).

Concluding Remarks

It should come as no surprise to educators that folk-biological knowledge dictates much of what students think they know about evolutionary relationships. We believe that

cladograms, when properly understood, provide powerful teaching tools that initiate discussion and challenge students to confront and replace their misconceptions with well-supported scientific explanations. The Study 3 results support this conjecture. The importance of helping students to overcome their misconceptions about biological classification cannot be overestimated: Evolutionary relationships sanction inferences that are critical in many domains, including human health, agriculture, and biotechnology (e.g., AMNH, 2002; Futuyma, 2004; Yates et al., 2004). For example, a recent public health crisis elucidates why it is essential to understand that fungi are more closely related to animals than to any other taxa. In the fall of 2012, more than 13,500 people in the United States received injections of a tainted steroid pain medication, leading to over 750 cases of fungal infections (including fungal meningitis) and the death of 64 patients (as of October 23, 2013; http://www.cdc.gov/hai/outbreaks/meningitis-map-large.html). The severity of this public health crisis is due to the difficulty of treating fungal, compared with bacterial and viral, infections. This is easy to appreciate when one understands that because fungi are the sister group to animals, sharing many cellular homologies, most antifungal drugs that negatively affect the pathogen also negatively affect the patient (Marcos, Gandia, Harries, Camona, & Munoz, 2012). Because long-standing misconceptions are resistant to "teaching away," we suggest that evolutionary taxonomy and tree thinking should be introduced to students at least by middle school.

APPENDIX: THE MATCHED PAIRS OF CLADOGRAMS AND ASSOCIATED QUESTIONS RECEIVED BY STUDENTS IN STUDY 1

Structure 1 Cladograms (see Figure 1)
* Unfamiliar taxa
  Answers for Question 1: Alejandro: *lampshade spider + crab spider + jumping spider;* Juan: *orchard spider + cave spider + comb-footed spider* [correct]; Carlos: *crab spider + jumping spider + orchard spider*
  Question 2: Given that orchard spiders have modified spigots, which other taxa (could be one or more) is/are most likely to share this character?
* Misconceptions taxa
  Answers for Question 1: Lauren: *mushroom + badger + fox* [correct]; Emily: *seaweed + grass + geranium;* Samantha: *grass + geranium + mushroom*
  Question 2: Given that mushrooms produce chitin, which other taxa (could be one or more) is/are most likely to share this character?

Structure 2 Cladograms (see Figure 2)
* Unfamiliar taxa
  Answers for Question 1: Lashonda: *Eurotiomycetes + Lichinomycetes + Sordariomycetes + Dothideomycetes + Arthoniomycetes* [correct]; Ebony: *Sordariomycetes + Dothideomycetes;* Nia: *Eurotiomycetes + Lichinomycetes + Sordariomycetes*
  Question 2: Given that *Dothideomycetes* produces a certain component of coenzyme Q, which other taxa (could be one or more) is/are most likely to share this character?
* Misconceptions taxa>
  Answers for Question 1: Saul: *dolphin + chimpanzee;* Aaron: *beaver + seal + dolphin + chimpanzee + bat* [correct]; Reuven: *beaver + seal + dolphin*
  Question 2: Given that chimpanzees have the epsilon-globin gene, which other taxa (could be one or more) is/are most likely to share this character?

Structure 3 Cladograms (see Figure 3)

* Unfamiliar taxa

  Answers for Question 1: Ilana: *S. uberis + S. pyogenes + S. canis + S. iniae;* Rachel: *S. phocae + S. agalactiae;* Tamar: *S. iniae + S. phocae + S. agalactiae + S. dysgalactiae* [correct]

  Question 2: Given that *S. iniae* has a certain nucleotide sequence on gene *rnpB*, which other taxa (could be one or more) is/are most likely to share this character?

* Misconceptions taxa

  Answers for Question 1: Malcolm: *shark + bass + salmon + lungfish;* Jamal: *lungfish + frog + salamander + moose* [correct]; Deshaun: *frog + salamander*

  Question 2: Given that lungfish have tooth enamel, which other taxa (could be one or more) is/are most likely to share this character?

Structure 4 Cladograms

Unfamiliar taxa: ((stone pine + (Turkish pine + Aleppo pine)) + ((Scots pine + red pine) + (Sikang pine + black pine)))

  Answers for Question 1: Matthew: *Turkish pine + Aleppo pine + Scots pine;* Andrew: *stone pine + Turkish pine + Aleppo pine* [correct]; William: *stone pine + Sikang pine + black pine*

  Question 2: Given that the Scots pine has a reduced number of resin canals, which other taxa (could be one or more) is/are most likely to share this character?

* Misconceptions taxa: ((bison + (porpoise + whale)) + ((manatee + elephant) + (horse + rhinoceros)))

  Answers for Question 1: Maria: *porpoise + whale + manatee;* Fernanda: *bison + horse + rhinoceros;* Gaudalupe: *bison + porpoise + whale* [correct]

  Question 2: Given that manatees have a circumferential placenta, which other taxa (could be one or more) is/are most likely to share this character?

Note: Cladograms preceded by an asterisk (*) were also used in Study 2.

We thank Emily Schreiber and Marissa Mencio for their help in collecting the data and Brenda Phillips for her help in coding students' written justifications.

REFERENCES

Ainsworth, S., & Loizou, A. Th. (2003). The effects of self-explaining when learning with text or diagrams. Cognitive Science, 27, 669–681.

Alverson, D. E., Smith, L. C., & Readence, J. E. (1985). Prior knowledge activation and the comprehension of compatible and incompatible text. Reading Research Quarterly, 20, 420–436.

American Museum of Natural History. (2002). Assembling the tree of life: Harnessing life's history to benefit science and society. Brochure produced for the National Science Foundation based on three NSF Tree of Life workshops held in 1999 and 2000 at Yale University, the University of California, Davis, and the University of Texas, Austin. Retrieved August 30, 2010, from http://www.phylo.org/sub_sections/outreach/outreach_b.php.

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars [target article]. Behavioral and Brain Sciences, 21, 547–609.

Barwise, J., & Etchemendy, J. (1991). Visual information and valid reasoning. In W. Zimmermann & S. Cunningham (Eds.), Visualization in teaching and learning mathematics (pp. 9–24). Washington, DC: Mathematical Association of America.

Baum, D. A., Smith, S. D., & Donovan, S. S. (2005). The tree thinking challenge. Science, 310, 979–980.

Berlin, B., Breedlove, D. E., & Raven, P. H. (1973). General principles of classification and nomenclature in folk biology. American Anthropologist, 75, 214–242.

Bowen, G. M., & Roth, W.-M. (2002). Why students may not learn to interpret scientific inscriptions. Research in Science Education, 32, 303–327.

Catley, K. M. (2006). Darwin's missing link: A new paradigm for evolution education. Science Education, 90, 767–783.

Catley, K. M., Lehrer, R., & Reiser, B. (2005). Tracing a prospective learning progression for developing understanding of evolution. Paper Commissioned by the National Academies Committee on Test Design for K-12 Science Achievement. Available from the first author.

Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. BioScience, 58, 976–987.

Catley, K. M., Novick, L. R., & Funk, D. J. (2012). The promise and challenges of introducing tree thinking into evolution education. In K. Rosengren, E. M. Evans, S. Brem, & G. Sinatra (Eds.), Evolution challenges: Integrating research and practice in teaching and learning about evolution (pp. 93–118). New York: Oxford University Press.

Catley, K. M., Phillips, B. C., & Novick, L. R. (2013). Snakes and eels and dogs! Oh, my! Evaluating high school students' tree-thinking skills: An entry point to understanding evolution. Research in Science Education, 43, 2327–2348.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. Review of Educational Research, 63, 1–49.

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. Journal of Research in Science Teaching, 35, 623–654.

Clement, J. (2008). The role of explanatory models in teaching for conceptual change. In S. Vosniadou (Ed.), International handbook of research on conceptual change (pp. 417–452). New York: Routledge.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Dufour-Janvier, B., Bednarz, N., & Belanger, M. (1987). Pedagogical considerations concerning the problem of representation. In C. Janvier (Ed.), Problems of representation in the teaching and learning of mathematics (pp. 109–122). Hillsdale, NJ: Erlbaum.

Freeman, S. (2011). Biological science: Evolution, diversity, & ecology (Vol. 2). Boston: Benjamin Cummings.

Futuyma, D. J. (2004). The fruit of the tree of life. In J. Cracraft & M. J. Donoghue (Eds.), Assembling the tree of life (pp. 25–39). New York: Oxford University Press.

Gilbert, S. F. (2003). Opening Darwin's black box: Teaching evolution through developmental genetics. Nature Reviews Genetics, 4, 735–741.

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. Psychological Science, 20, 480–487.

Goldsmith, D. W. (2003). Presenting cladistic thinking to biology majors & general science students. The American Biology Teacher, 65, 679-682.

Gregory, T. R. (2008). Understanding evolutionary trees. Evolution: Education and Outreach, 1, 121–137.

Halverson, K. L., Pires, C. J., & Abell, S. K. (2011). Exploring the complexity of tree thinking expertise in an undergraduate systematics course. Science Education, 95, 794–823.

Hampton, J. A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 12–32.

Hegarty, M., Carpenter, P. A., & Just, M. A. (1991). Diagrams in the comprehension of scientific texts. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research (Vol. 2, pp. 641–668). New York: Longman.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. Journal of Memory and Language, 32, 717–742.

Hegarty, M., & Stull, A. T. (2012). Visuospatial thinking. In K. J. Holyoak & R. G. Morrison (Eds.), Oxford handbook of thinking and reasoning (pp. 606–630). New York: Oxford University Press.

Hennig, W. (1966). Phylogenetic systematics. Urbana: University of Illinois Press.

Johnson, K. E., Mervis, C. B., & Boster, J. S. (1992). Developmental changes within the structure of the mammal domain. Developmental Psychology, 28, 74–83.

Kendeou, P., & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. Journal of Educational Psychology, 97, 235–245.

Kindfield, A. C. H. (1993/1994). Biology diagrams: Tools to think with. Journal of the Learning Sciences, 3, 1–36.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11, 65–99.

Lee, M. S. Y., Reeder, T. W., Slowinski, J. B., & Lawson, R. (2004). Resolving reptile relationships: Molecular and morphological markers. In J. Cracraft & M. J. Donoghue (Eds.), Assembling the tree of life (pp. 451–467). New York: Oxford University Press.

Life Science. (2007). Holt McDougal.

Lynch, M. (1990). The externalized retina: Selection and mathematization in the visual documentation of objects in the life sciences. In M. Lynch & S. Woolgar (Eds.), Representation in scientific practice (pp. 153–186). Cambridge, MA: MIT Press.

Maienschein, J. (1991). From presentation to representation in E. B. Wilson's The Cell. Biology and Philosophy, 6, 227–254.

Marcos, J. F., Gandia, M., Harries, E., Carmona, L., & Munoz, A. (2012). Antifungal peptides: Exploiting non-lytic mechanisms and cell penetration properties. In K. Rajasekaran, J. W. Cary, J. M. Jaynes, & E. Montesinos (Eds.), Small wonders: Peptides for disease control (ACS Symposium Series, Vol. 1095, pp. 337–357). Washington, DC: American Chemical Society.

McKim, R. H. (1980). Thinking visually: A strategy manual for problem solving. Belmont, CA: Wadsworth.

Meir, E., Perry, J., Herron, J. C., & Kingsolver, J. (2007). College students' misconceptions about evolutionary trees. The American Biology Teacher Online, 69 (7), e71–e76.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. Bulletin of the Psychonomic Society, 7, 283–284.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K. Google Books Team, et al. (2011). Quantitative analysis of culture using millions of digitized books. Science, 331, 176—182.

Miller, K. R., & Levine, J. (2002). Biology. Upper Saddle River, NJ: Prentice Hall.

Morabito, N., Catley, K. M., & Novick, L. R. (2010). Reasoning about evolutionary history: The effects of biology background on post-secondary students' knowledge of most recent common ancestry and homoplasy. Journal of Biological Education, 44, 166–174.

Nehm, R. H., Beggrow, E. P., Opfer, J. E., & Ha, M. (2012). Reasoning about natural selection: Diagnosing contextual competency using the ACORNS instrument. The American Biology Teacher, 74, 92–98.

Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. Journal of Research in Science Teaching, 48, 237–256.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 54–65.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 510–520.

Novick, L. R. (2003). At the forefront of thought: The effect of media exposure on airplane typicality. Psychonomic Bulletin & Review, 10, 971–974.

Novick, L. R. (2006). The importance of both diagrammatic conventions and domain-specific knowledge for diagram literacy in science: The hierarchy as an illustrative case. In D. Barker-Plummer, R. Cox, & N. Swoboda (Eds.), Diagrammatic representation and inference (Diagrams 2006). Lecture notes in artificial intelligence (Vol. 4045, pp. 1–11). Berlin: Springer-Verlag.

Novick, L. R., & Catley, K. M. (2013). Reasoning about evolution's grand patterns: College students' understanding of the Tree of Life. American Educational Research Journal, 50, 138–177.

Novick, L. R., Catley, K. M., & Funk, D. J. (2011). Inference is bliss: Using evolutionary relationship to guide categorical inferences. Cognitive Science, 35, 712–743.

Novick, L. R., Schreiber, E. G., & Catley, K. M. (2014). Fostering 21st century evolutionary reasoning in undergraduates: A research-based approach to teaching tree thinking. Manuscript submitted for publication.

O'Hara, R. J. (1988). Homage to Clio, or, toward an historical philosophy for evolutionary biology. Systematic Zoology, 37, 142–155.

Opfer, J. E., Nehm, R. H., & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. Journal of Research in Science Teaching, 49, 744–777.

Osherson, D. N., Smith, E. E., Wilkie, O., & López, A. (1990). Category-based induction. Psychological Review, 97, 185–200.

Phillips, B. C., Novick, L. R., Catley, K. M., & Funk, D. J. (2012). Teaching tree thinking to college students: It's not as easy as you think. Evolution: Education and Outreach, 5, 595–602.

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2011). Campbell biology (9th ed.). San Francisco: Benjamin Cummings.

Rotbain, Y., Marbach-Ad, G., & Stavy, R. (2006). Effect of bead and illustrations models on high school students' achievement in molecular genetics. Journal of Research in Science Teaching, 43, 500–529.

Sandvik, H. (2008). Tree thinking cannot taken for granted: Challenges for teaching phylogenetics. Theory in Biosciences, 127, 45–51. Retrieved March 19, 2008, from http://www.springerlink.com/content/eu62420p381402xr/.

Shapiro, S. I., & Palermo, D. S. (1970). Conceptual organization and class membership: Normative data for representatives of 100 categories. Psychonomic Monograph Supplements, 3(11, Whole No. 43), 107–127.

Stavy, R. (1991). Using analogy to overcome misconceptions about conservation of matter. Journal of Research in Science Teaching, 28, 305–313.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. Journal of Experimental Psychology: General, 119, 176–192.

Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: Understanding the wrong answers—Teaching toward conceptual change. Cell Biology Education, 4, 112–117.
Thanukos, A. (2009). A name by any other tree. Evolution: Education & Outreach, 2, 303–309.
Trowbridge, J. E., & Mintzes, J. J. (1988). Alternative conceptions in animal classification: A cross-age study. Journal of Research in Science Teaching, 25, 547–571.
Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In D. L. Gabel (Ed.), Handbook of research on science teaching and learning (pp. 177–210). New York: Macmillan.
Winn, W. (1989). The design and use of instructional graphics. In H. Mandl & J. R. Levin (Eds.), Knowledge acquisition from text and pictures (pp. 125–144). Amsterdam, The Netherlands: Elsevier.
Yates, T. L., Salazar-Bravo, J., & Dragoo, J. W. (2004). The importance of the tree of life to society. In J. Cracraft & M. J. Donoghue (Eds.), Assembling the tree of life (pp. 7–17). New York: Oxford University Press.
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. Developmental Review, 27, 172–223.

Science
Education

LEARNING

An Integrative Framework for the Analysis of Multiple and Multimodal Representations for Meaning-Making in Science Education

KOK-SING TANG,[1] CESAR DELGADO,[2] ELIZABETH BIRR MOJE[3]
[1]*National Institute of Education, Nanyang Technological University, Singapore 637616;* [2]*College of Education, University of Texas at Austin, Austin, TX 78712, USA;* [3]*School of Education, University of Michigan, Ann Arbor, MI 48109, USA*

Received 30 August 2012; accepted 18 December 2013
DOI 10.1002/sce.21099
Published online 30 January 2014 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: This paper presents an integrative framework for analyzing science meaning-making with representations. It integrates the research on multiple representations and multimodal representations by identifying and leveraging the differences in their units of analysis in two dimensions: timescale and compositional grain size. Timescale considers the duration of time a learner typically spends on one or more representations. Compositional grain size refers to the elements of interest within a representation, ranging from components such as visual elements, words, or symbols, to a representation as a whole. Research on multiple representations focuses on the practice of re-representing science concepts through different representations and is typically of long timescale and large grain size. Research on multimodal representations tends to consider how learners integrate the components of a representation to produce meaning; it is usually of finer grain size and shorter timescale. In the integrative framework, each type of analysis on multiple and multimodal representations plays a mutually complementary role in illuminating students' learning with representations. The framework is illustrated through the analysis of instructional episodes of middle school students using representations to learn nanoscience concepts over the course of a lesson unit. Finally, recommendations for new research directions stemming from this framework are presented. © 2014 Wiley Periodicals, Inc. *Sci Ed* **98:**305–326, 2014

Correspondence to: Kok-Sing Tang; e-mail: koksing.tang@nie.edu.sg
This research is supported by the National Center for Learning and Teaching in Nanoscale Science and Engineering (NCLT) under National Science Foundation Grant #0426328.

INTRODUCTION

Representations are artifacts that symbolize an idea or concept in science (e.g., force, energy, chemical bonding) and can take the form of analogies, verbal explanations, written texts, diagrams, graphs, and simulations. As such, they are an integral part of the language of science. The National Science Foundation (NSF), recognizing the need for a greater understanding of representation, funded two "cross-border" conferences that brought together researchers from the literacy, cognitive science, and science education communities. These conferences determined that further syntheses and frameworks are needed to explain how representation promotes science literacy (Hand et al., 2003). Specifically, greater understanding is needed on two areas of research on representation: multiple representations and multimodal representations (Yore & Treagust, 2006).

The term "multiple representations" denotes the practice of representing to students the same concept through different representational forms (Prain & Waldrip, 2006). Research on *multiple representations* has focused on how the use of more than one representation affects student understanding (e.g., Ainsworth, 2006; Gilbert & Treagust, 2009; Kozma, 2003; Prain, Tytler, & Peterson; 2009). The term "multimodal representations" refers to the fact that learning with one or more representations usually integrates components of various modalities such as language, depiction, and symbols (Prain & Waldrip, 2006). This area of research on *multimodality* examines how students build scientific understanding through the simultaneous use of various modalities within and across representations (e.g., Airey & Linder, 2009; Kress, Jewitt, Ogborn, & Tsatsarelis, 2001; Lemke, 1998).

Research on multiple representations and multimodality are well established in science education research. However, there have been few attempts to integrate these disparate areas of research (Yore & Hand, 2010). We posit that the central difficulty in linking the two research areas lies in their different units of analysis, focusing on the number of representations used in a teaching or learning context. The different units of analysis result in differences along two dimensions: timescale and compositional grain size. The purpose of this paper is to present a framework that leverages these two dimensions to connect the two areas of research, as well as to suggest additional directions for research. In doing so, this paper advances the vision of a multirepresentational framework first put forth by Yore and Treagust (2006).

In the next sections, we define the two dimensions of timescale and compositional grain size and use them to organize prior research on representation. We then present our framework and show how it integrates multiple representations and multimodal foci through an analytical case study. Finally, the implications of this framework for future research on representation are discussed.

THEORETICAL BACKGROUND

Timescale and Compositional Grain Size as Dimensions of the Unit of Analysis

As mentioned above, research on multiple representations and multimodality typically use different units of analysis. In studying how several representations can interact to support student learning, research on multiple representations usually considers a longer timescale and uses a larger compositional grain size. On the other hand, research on multimodality examines how learners make sense of a representation consisting

of multiple modalities and is characterized by shorter timescales and finer compositional grain size.

Timescale. According to Lemke (2000), there are characteristic timescales of reoccurring processes observed in classroom events, ranging from a single utterance in seconds, to an exchange of teacher–student or student–student dialogue in minutes, a full lesson in an hour, a lesson unit in days, and finally a curriculum and program in months and years. To understand classroom events, one must observe how the processes at a shorter timescale build up to the processes at a longer timescale and conversely how the longer timescale processes constrain and enable the kind of processes that can occur at a shorter timescale (Lemke, 2000). Making sense of the elements within a representation, which is the focus of multimodality, usually involves shorter temporal scales of seconds or a few minutes. Using and transforming several tables, diagrams, or graphs from one form to another, which is the focus of multiple representations, usually involves longer timescale of at least one lesson period. The dimension of timescale is continuous, but it is conceptually useful to divide it into two levels. For the purpose of studying representation, we define a short timescale as less than a lesson period.

Compositional Grain Size. Compositional grain size refers to the elements that make up a representation (Tang & Moje, 2010). For a written text, compositional grain sizes could range from letters as the smallest components, to words, phrases, clauses, sentences, paragraphs, pages, and sections. For a visual diagram, the components could range from lines or shapes to the entire diagram. For example, the finest grained components of a molecular depiction of air exerting pressure on its container are the lines (representing the container), dots (representing molecules), and arrows (representing motion) drawn in the diagram. The intermediate components of a diagram are clusters, or local groupings of spatially proximate items, which define a specific subregion of the diagram as a whole (see Baldry & Thibault, 2006). Finally, the largest compositional grain size can be one or more diagrams in their entirety. While many compositional grain sizes can be defined, we define just two levels, with fine grain size defined as consisting of less than an entire representation.

Relationship Between Timescale and Compositional Grain Size

Archetypically, multiple representation studies feature longer timescale and larger compositional grain size, whereas multimodality studies are characterized by shorter timescale and finer compositional grain size. However, the two dimensions of timescale and grain size are independent, and thus define four possible combinations. We depict these in a two-by-two space with timescale on the horizontal dimension and grain size on the vertical dimension. We next describe representative studies that fall in each of the four quadrants. Figure 1 shows this two-by-two space, populated by research studies in each of the quadrants.

Analysis With Long Timescale and Large Grain Size. A good example of an analysis of multiple representations with a long timescale and large grain size (the top left quadrant of Figure 1) is the study by Hubber, Tytler, and Haslam (2010) in the context of forces. They focused not only on the representations used in class but also on "re-representation": How representations were transformed from one representation to the next (e.g., drawing

Grain size \ Timescale	Long timescale	Short timescale
Large grain size	Hand, Gunel, & Ulu (2009) Hubber, Tytler, & Haslam (2010) Kozma, Chin, Russell, & Marx (2000) Prain, Tytler, & Peterson (2009) Roth & Tobin (1997) Waldrip, Prain, & Carolan (2010)	Ainsworth, Bibby, & Wood (2002) Mayer (2001) Rosengrant, Etkina, & Van Heuvelen (2007) Schnotz & Bannert (2003)
Fine grain size	Jaipal (2009) Kress, Jewitt, Ogborn & Tsatsarelis (2001) Márquez, Izquierdo, & Espinet (2006)	Fredlund, Airey, & Linder (2012) Lemke (1998) Pozzer-Ardenghi & Roth (2007) Tang, Tan, & Yeo (2011)

Figure 1. Map of the problem space: timescale and grain size. Studies are classified according to their temporal and compositional characteristics.

to table to graph). Re-representation can occur within the same modality (e.g., from one written text to another) or across multiple modalities (e.g., from text to graphs). Working at a timescale over 12 lessons, they worked with the teachers to develop what they called "a representational approach" to the teaching of forces. The pedagogical principles in this approach include (i) introducing multiple representations of the concept, (ii) encouraging students to generate their own representations, and (iii) linking representations to experiential activity, discussion, cognition, and communication. In one of their analyses, they studied how the dynamic transformations in the students' representations corresponded with the teaching and learning sequences. The sequence of re-representations made by the students included everyday words associated with forces, gestures miming the actions, drawings of their actions, drawings of the effects on modeling clay, and force diagrams with arrows. The unit of analysis selected by Hubber and colleagues comprised multiple representations used in a lesson, because they were interested in re-representation. This focus required a longer timescale to study the dynamic generation and negotiation of representations and their transformations. The large grain size corresponds to a view of a representation as a self-contained artifact designed with some specific science concept in mind. Additional studies of this type are listed in Figure 1.

Analysis With Short Timescale and Fine Grain Size. An example of an analysis with a finer grain size and shorter timescale (at bottom right quadrant of Figure 1) is Lemke's (1998) multimodal analysis of printed scientific texts. In one of his analytical examples, Lemke first decomposed a figure into various visual components such as shaded circles, arrow vectors, parallel lines, and dashed lines. He then considered how the qualities of each component relate to those of other components in the construction of scientific meanings. Lemke did not observe readers interacting with the representation but presumably this interaction would occur over a period of a few minutes. In another multimodal study, Tang, Tan, and Yeo (2011) analyzed the critical connections among multimodal elements that constitute the concept of work-energy. They analyzed in detail three episodes of the discussion among a group of students, each lasting a few minutes. They

found that students constructed knowledge through the integration of four modalities: language, diagrams, mathematical symbolism, and gestures. Each modality had different roles and functions. Both studies feature a unit of analysis of a single representation, observed over a shorter timescale, and focusing on the fine-grained compositional elements of various modalities that constitute the representation. Their ultimate aim was to analyze how those elements related with one another in the overall construction of scientific concepts.

These and other multimodal studies are based on Halliday's (1978) theory of social semiotics. Social semiotics is the study of sign systems and their use in meaning-making as a function of a social process. An important notion in social semiotics is *semiotic affordances*, which examines the possibility of different kinds of meaning that are made available through the use of different modalities (Kress et al., 2001). For instance, a linguistic modality in general allows or affords a person to make categorical types of meaning (e.g., of what kind), whereas a visual modality affords a person to make quantitative types of meaning (e.g., by how much). Multimodality and semiotic affordances are useful notions because they provide a metalanguage and analytical tools to examine the fine-grained components of a representation and to understand how the components come together to form meanings (see Figure 1 for additional studies with short timescale and fine grain size).

Analysis With Short Timescale and Large Grain Size. Studies in this category (top right in Figure 1) analyze the design features and parameters of different representations used to promote conceptual learning. Ainsworth (2006) studied the different functions of multiple representations and generated a taxonomy of functions that included constraining interpretations, complementing each other, or constructing deeper meaning. Schnotz and Bannert (2003) studied how learners use text and pictures to construct their understanding. Drawing from Chandler and Sweller's (1996) cognitive load theory and Mayer's (2001) dual sensory processing theory, they proposed an integrated model of text and picture comprehension. Based on this model, they designed a randomized-trial experiment to compare learning with text alone and with text and diagrams of two different types. The unit of analysis in both studies is one or more entire representations. The timescale associated with these studies is short, at the level of a task (ranging from 1 to 5 minutes). The compositional grain size is large, composed of representations as a whole. Other studies with short timescale and large grain size are listed in Figure 1.

Analysis With Long Timescale and Fine Grain Size. Studies of this type (bottom left quadrant in Figure 1) use a fine-grained analytical approach, but investigate a phenomenon that occurs at a longer timescale. Márquez and colleagues (2006) were interested in the communicative roles of different modalities used by a secondary science teacher. They studied a lesson unit on the water cycle, composed of five 55-minutes lessons. They used Halliday's (1994) linguistic framework and Kress and van Leeuwen's (2006) visual framework to carry out a fine-grained decomposition of the verbal discourse and pictorial representations, respectively. For instance, in analyzing a visual representation of the water cycle, Márquez et al. examined the arrows within this representation and identified three different meanings of these arrows within the context of the visual representation. They then used categorization and statistical analyses to investigate the functions of speech, gesture, and diagram in relation to the thematic construction of water cycle. Although various representations were involved, the interactions of the components of each representation were considered, making this a fine-grained analysis with a long timescale. Additional studies with long timescale and fine grain size are shown in Figure 1.

AN INTEGRATIVE FRAMEWORK FOR THE ANALYSIS OF MULTIPLE AND MULTIMODAL REPRESENTATIONS

From the above analysis of the literature in representation studies, one can see the disparate foci based on the two dimensions of timescale and compositional grain size. We developed our framework with the aim of incorporating a wider range of timescale and grain size in the analysis of representation, in effect integrating across multiple representations and multimodal approaches.

We begin developing our framework from the definition of a representation as a designed artifact. Drawing from the literature on multiple representations, we incorporate the theoretical notion of re-representation (Hubber et al., 2010) as the transformation of representations from one artifact to another across a continuous chain of human activities. This expands our scope from representation as an artifact to representation as a process of meaning-making that makes use of representations as mediating tools. This also broadens the timescale of analysis from an activity involving one designed artifact (usually in minutes) to a sequence of representational activities in a lesson or lesson unit. We then draw on the literature on multimodality to incorporate the notion of semiotic affordances, which examines the possibilities and constraints of a representation's meaning-making potential. A focus on semiotic affordances expands the range of our compositional grain size from a representation as a unitary whole to include the smaller semiotic elements that constitute the artifact and its meaning potential. The focus on semiotic affordances allows us to examine how the short timescale events build up to the processes at a longer timescale, whereas the incorporation of re-representation affords understanding how the longer timescale events constrain and support shorter timescale events (Lemke, 2000).

Our integrative framework is shown visually in Figure 2. The notion of re-representations (at top of Figure 2) considers the sequences of representations that might be used in a lesson unit, focusing on the process of transforming one representation to the next and also how one representation relates to the others (e.g., constraining interpretations; Ainsworth, 2006). In this example, the naked-eye examination of a sample is followed by the use of a microscope, with students producing drawings of each (the first two objects from left to right). Students next produce a diagram (third object from left) that captures important qualitative aspects of the phenomenon, then produce measurements that they organize into a table, then a graph displaying the mathematical equation that models the phenomenon (the last two objects). This level of analysis involves a longer timescale and larger compositional grain size. On the other hand, the semiotic affordances analysis (at bottom of Figure 2) takes a fine-grained look at one representation at a time. It examines how the composition and integration of the various elements (lines, curves, arrows, boxes, words, symbols, numbers) afford a person who is using it to construct meaning related to the phenomenon. This process usually occurs at a shorter timescale. The relationship between re-representation and semiotic affordances (the top and bottom part of Figure 2) is iterative and cyclical, each analysis informing the other. In the next section, we demonstrate the use of this framework through a case study.

METHODS

Research Context and Data Sources

We use a case study to illustrate our integrative framework, which we originally developed to understand how student make meanings with representations. Our case study is located in a free 2-week summer program in the U.S. Midwest, attended by 40 students from

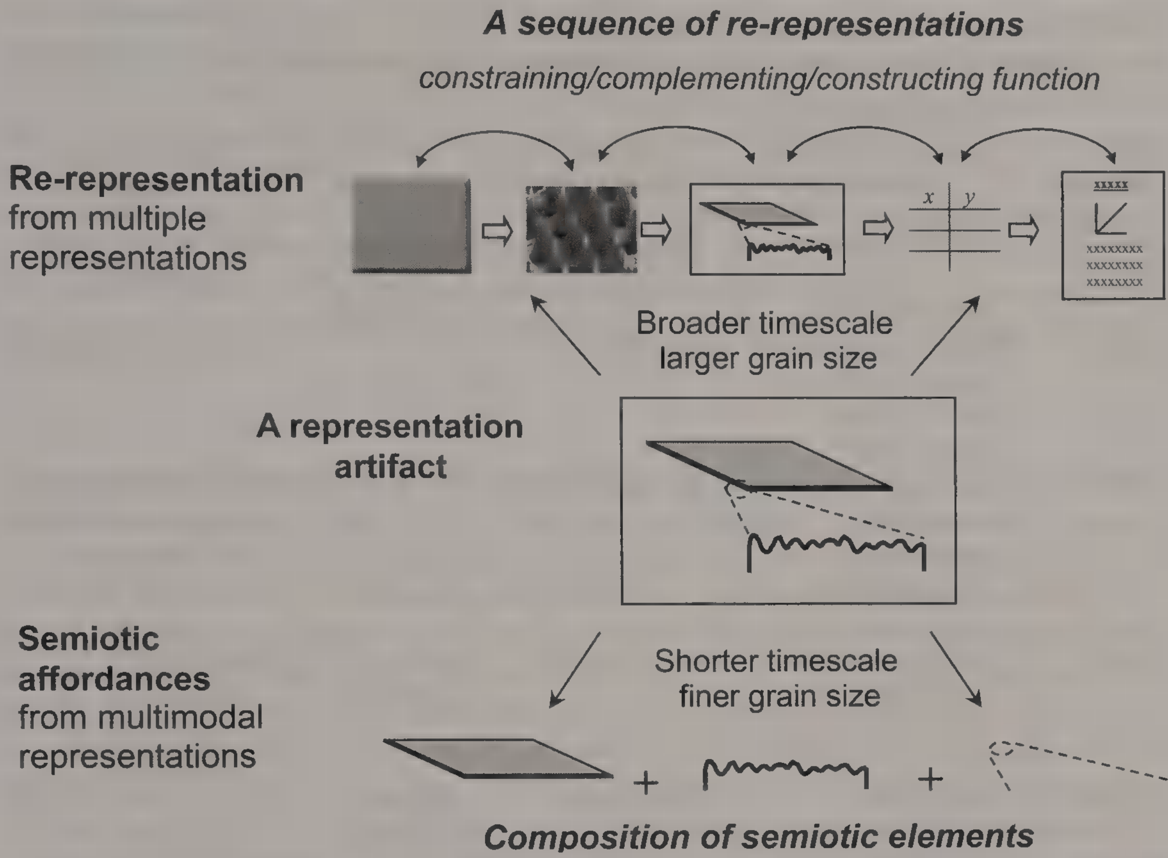


Figure 2. Our integrative framework for the analysis of multiple and multimodal representations.

two local middle schools who volunteered to participate. The summer program was a collaboration between an NSF-funded research center, the school district, and an outreach program affiliated with a university hospital. We researched a curriculum strand designed to teach the concepts of size and scale in six lessons.

The curriculum was designed following a project-based pedagogical approach (Krajcik & Blumenfeld, 2006) where the students learn by carrying out a series of empirical investigations to address a real-world problem. The project was contextualized with the real-life case of a middle school student who died of an antibiotic-resistant bacterial infection contracted at his school. Students worked in groups to investigate the problem, create artifacts, and report their suggested solutions. Our use of representations was guided by studies on multiple representations. Representations used in the lesson activities included physical, three-dimensional scale models of nanoscale objects such as DNA and viruses along with two- and one-dimensional representations of larger objects (e.g., cells) at the same scale, videos of commercial products designed to reduce bacterial infections, and computer visualizations. Consistent with Hubber et al.'s (2010) re-representational approach, students generated their own representations linked to various activities throughout the lessons, such as sketches of objects viewed under the microscope and posters presenting their ideas.

The primary data source was videotaped observations, recorded by one camera focused on the lead teacher and another focused on the interactions of several groups of students. In this paper, we report our observations primarily from a group of students consisting of Mary, Luke, and Dave (all names are pseudonyms to protect privacy), with the other groups providing confirming and disconfirming evidences to our assertions. Additional data sources were observation fieldnotes, instructional materials, and students' completed artifacts. The first author took the role of a participant-observer in collecting the data.

TABLE 1
An Example of Segmentation and Corresponding Tags from the First Lesson

Video Time	Description of Teaching/Learning Activities	Participation Structure	Thematic Content	Representations Used
1:33:34	Teacher introduces sandpaper experiment	Monologue + lab demo	Modeling bacteria buildup	Sandpapers + salt (apparatus)
1:35:45	Teacher instructs students on how to report on worksheet	Monologue		Table on whiteboard
1:36:16	Individual students carry out experiment and record on their worksheet	Seatwork		Sandpapers + salt (apparatus), table on worksheet
1:44:26	Teacher draws and explains what a side view is	Class dialogue	Surface feature	Sandpaper (apparatus), drawing on whiteboard
1:45:46	Groups draw a poster to explain the sandpaper experiment	Group work	Explaining bacteria buildup	Sandpapers + salt (apparatus), writing/drawings on poster paper
2:03:47	Teacher gives instruction for presentation	Monologue		Nil
2:04:25	First group presents	Group presentation		Writing/drawings on poster paper

Data Analysis

Initial Analysis: Re-Representation. Our initial analysis focused on the multiple representations used in the six lessons. Each representation is analyzed as a whole, with a long timescale and large compositional grain size. Lesson videos were viewed, coded, and tagged using Transana software. We first segmented the data by dividing the continuous sequence in a lesson video into meaningful discrete units. The average time of a segment is 4.5 minutes. We coded and tagged each segment according to four categories: teaching activities (e.g., teacher explanation or group experiment), participation structures (e.g., teacher monologue or class discussion), thematic content (e.g., bacteria buildup), and the representations used (e.g., group poster); see Table 1. At this level of analysis, verbal dialogue from the video data was analyzed at the grain size of an exchange (a string of utterances between participants for a specific purpose). The dialogue was not transcribed at this point due to the time-consuming nature of transcription.

The tags inserted into the video allowed us to track the use of a particular representation throughout the lesson unit and follow how it was transformed by the teacher or students. In other words, we tracked the sequences of re-representations. This analysis provided insight into to the social process of learning with representations, identified the sequence

of re-representations involved, and allowed us to select focal representations for the next phase of analysis. This multiple representations analysis was consistent with our theoretical approach in designing the curriculum and yielded information on re-representation. However, we found it insufficient to fully explicate our observations. In particular, we wished to understand why different groups came up with varying interpretations of the same phenomena. We realized that we would need to delve deeper into how students interacted with a representation to make meaning. Only by examining the shorter timescale processes in a fine-grained manner could we understand the long timescale sequences of re-representation and learning across multiple lessons. This is what motivated us to use the analytical tools of multimodality.

Second-Phase Analysis: Semiotic Affordances. Our second-phase analysis followed a multimodal approach in focusing on a selected representation. With a short timescale and fine compositional grain size, we focused on the components of the analyzed representation. We selected two representations for further analysis based on our earlier analysis of the re-representation sequences. One representation was the result of the students' group discussions on the first lesson, and another was the product of their group presentations on the sixth and final lesson. These representations were selected first because of the multimodal richness of the corresponding episodes and second because the episodes on the first and final lesson could give a sense of the trajectory of the students' development of ideas over the lesson unit. In terms of the thematic content, the two representations dealt with "self-cleaning nanotech surfaces" (henceforth, *self-cleaning*). Surfaces that are smooth at the nanoscale harbor fewer bacteria and are used in commercial products including toilets (while electrostatics also influence the propensity of bacteria to cling to a surface, the curriculum only focused on surface roughness). Self-cleaning was being investigated in the context of the project-based unit on avoiding bacterial outbreaks at school.

For each of the selected representations, we transcribed the corresponding video segments and carried out a detailed multimodal discourse analysis. At this level of analysis, spoken language was analyzed at the level of a clause. Clauses function in English as the basic unit that semantically constructs a particular event or sense of experience (Halliday, 1978). Sentences may contain several clauses, joined together through conjunctions such as "because" or "and." We then interpreted the meaning of each clause through the semantic relationship among the words in the clause; for instance, the clause "the surface is bumpy" is an attributive relationship between a carrier and its attribute whereas "surface has bumps" is a possessive relationship between a carrier and its possessions. "You can feel the bumps" involves an agent—"you"—doing something to an object. (For a list of semantic relationships, see Lemke, 1990.)

For every verbal clause that we analyzed, we also examined the corresponding nonverbal actions and representations that the participants were oriented to in the video segment. Visual elements found in the representations were analyzed using Kress and van Leeuwen's (2006) visual framework. For instance, a common visual representation drawn by the students during the first lesson is called an analytical structure, which relates visual elements in terms of a part–whole structure between a carrier (the whole) and its possessive attributes (the parts). Nonverbal actions such as pointing gestures were used to determine the component(s) of a representation that a student was referring to, whereas iconic and metaphorical gestures often supplemented the verbal communication with further information (see McNeill, 2005 for the various types and functions of gestures). Examples of these analytical methods will be further illustrated in the analysis.

Iterative Nature of the Analysis. The findings from the semiotic affordances analysis were used to better understand a particular representation within a sequence of re-representation, which also shed light on the sequence as a whole. Likewise, prior and subsequent representations in the sequence helped us understand the fine-grained way in which students constructed meaning from one particular representation.

RESULTS

Initial Analysis: Re-Representation

The first lesson was aimed at building an understanding of self-cleaning. Students explored and modeled the role of surface roughness in allowing bacteria to cling onto a surface. The first activity was an experiment that used different grades of sandpaper to model surfaces of varying degrees of smoothness and grains of salt to model bacteria. Students explored the different degree of difficulty in removing the salt from each grade of sandpaper, using a note pad as a scraper. Subsequently, students used multiple representations (e.g., diagram, table) to construct scientific explanations of self-cleaning based on their observations. Table 2 shows the representations and learning activities used in this lesson.

According to our framework, these representations are artifacts designed to teach a science concept. While meaningful for the designer, they are initially devoid of any meaning to others—they are just sandpaper, salt, and a collection of writings and drawings. Meaning is made through the use of multiple representations. Each representation forms a part of a sequence of re-representations, and any meaning made with one representation depends on prior meanings made with preceding representations across space and time.

At the beginning of the lesson, Mary, Luke, and Dave experimented with the salt and sandpaper model. They individually recorded their observations in the form of written texts and drawings in a table in their individual worksheets. Next, they collaborated to prepare a group poster, which was to be used in a subsequent oral presentation, to explain their findings (see Figure 3). The sequence of re-representation included a physical experience transformed to a textual description and drawing, and then to a group poster. Following Hubber et al. (2010), we next analyze how the dynamic transformations in the students' representations corresponded with the teaching and learning sequences.

Initially, when the group started working on the group poster, with Luke drawing and Mary and Dave helping, they drew only the *top view* of the sandpaper to represent what they saw from the top looking down at the sandpaper (the rectangular images directly below the text labeling the three grades of sandpaper). About 5 minutes later, Mary interrupted Luke

TABLE 2
Representations Used in the Curricular Lesson

	Representation	Curricular Purpose
a	Sandpaper and grains of salts	To simulate different surface textures and bacteria respectively
b	Written table in a worksheet	For individual students to record their observations and explanations
c	Written text and a diagram in a worksheet	To describe to students the experimental procedures
d	Drawings on a shared poster paper	For group of students to present to the class

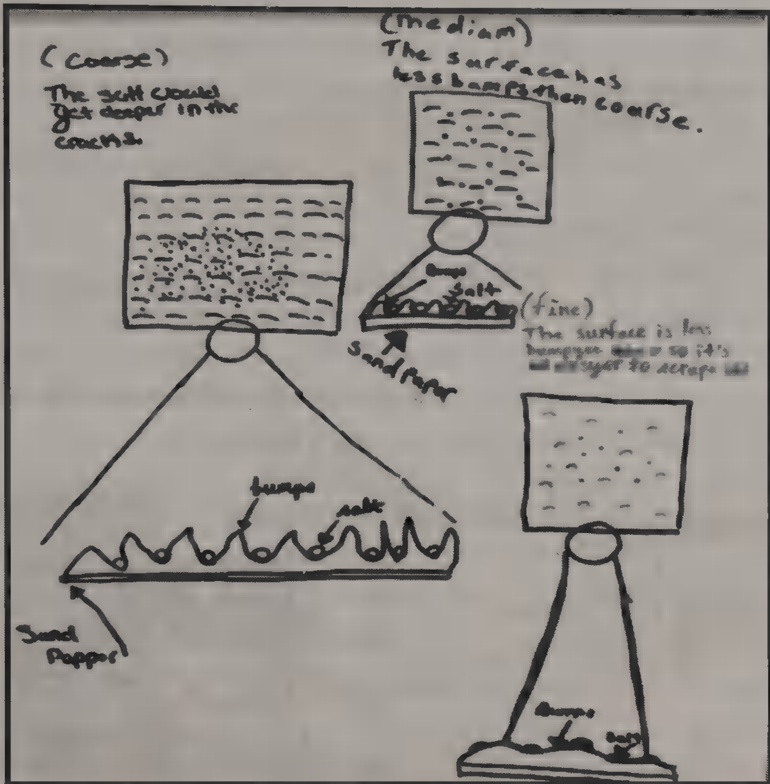


Figure 3. Text and drawings on group poster.

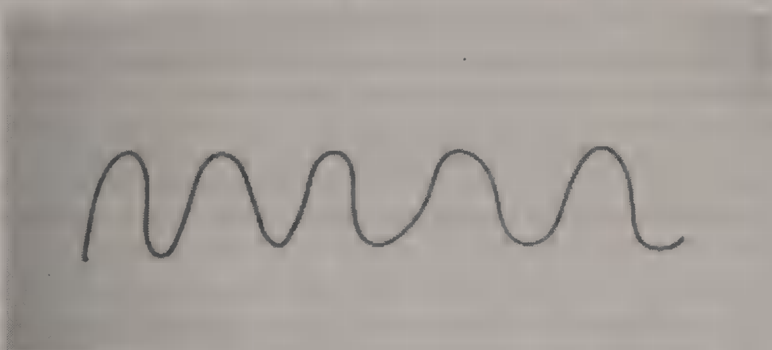


Figure 4. Drawing made by the teacher on the whiteboard.

and pointed out to her group an image made earlier by the lead teacher on the whiteboard (see Figure 4). The teacher made that image to explain the meaning of a *side view* before the class started working on the posters. Mary explained to the group that they needed to draw a side-view image of the sandpaper. She then turned to look at a diagram accompanying the written instruction on the worksheets (see Figure 5), took over the pen from Luke and proceeded to draw a "magnified" side-view image directly below the top-view drawing (labeled "sand paper" in Figure 3 at lower left).

Soon after Mary explained to the group what she was doing, Luke and Dave followed suit and each drew one side-view image extending from the top-view drawings on the group poster (Figure 3). About 3 minutes later, after the students had completed the poster, the researcher as the participant observer asked them to explain what they had drawn. Mary responded while looking and pointing at the top-view drawings on the poster:

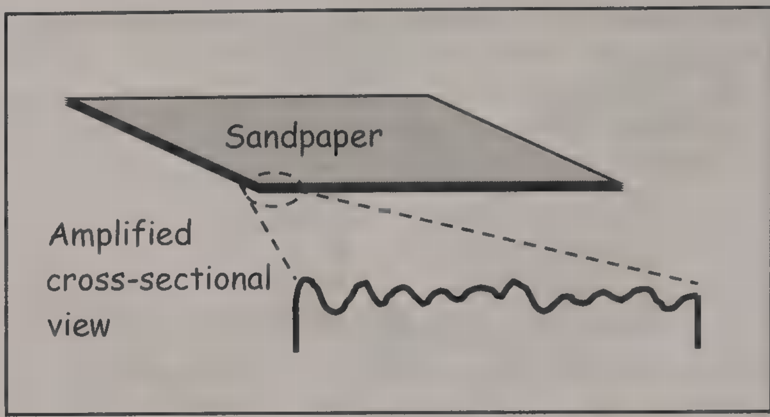


Figure 5. Diagram printed on a page of the student worksheets.

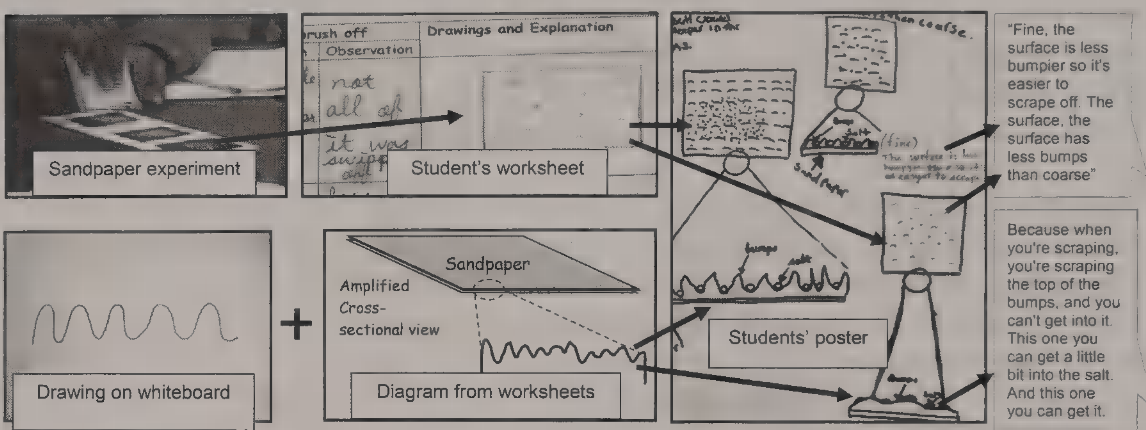


Figure 6. Sequences of re-representations leading to the students' group poster and their explanations of surface features. Top sequence shows the re-representation process from the sandpaper experiment to written text and top view drawings in each student's worksheet, and to top view drawings in the group poster. Bottom sequence shows how the teacher's drawing on the whiteboard, reproduced from the student worksheet (with accompanying text about it representing a magnified section of a side view), was re-represented to the side view drawings in the group poster, along with an elaboration of the magnified section of a side view (see text for explanation of the sequences).

Researcher: Ok, now that you have drawn all three, can you compare what you have drawn? And can you explain why this is the salt is easier to come out, for this. It is harder to come out?

Mary: For fine, the surface is less bumpier so it's easier to scrape off. The surface, the surface has less bumps than coarse.

Shortly after, the researcher pointed at the side-view drawings and posed the question again. Interestingly, the answer from Mary became very different. As she looked and pointed at the side-view drawings, she gave the following response:

Mary: Because when you're scraping, you're scraping the top of the bumps, and you can't get into it. This one you can get a little bit into the salt. And this one you can get it.

A summary of the sequences of re-representations leading to these two different responses is shown in Figure 6. As shown in the top sequence of Figure 6, the top views were re-represented from their drawings in the individual worksheets, which were themselves re-represented from the sandpaper experiment. In this sequence, the students'

top-view representations reflected their experiences in brushing off the salt from the sandpaper in the prior experimental activity. This resulted in Mary's explanation that the fine-grade sandpaper is easier to scrape off because it "is less bumpier" and "has less bumps."

By contrast, as shown in the bottom sequence of Figure 6, the side-view drawings in the group poster were re-represented from the diagram on the students' worksheets and the lead teacher's drawing on the whiteboard. This resulted in Mary's explanation that "you can't get into" the salt, when it is on the coarse sandpaper.

The pattern of different explanations for side-view and top-view representations held for other groups as well. Consider two other examples from groups that only considered the top views of the sandpaper in their presentations:

Adrian: On the fine piece of um (points at top-view drawing), whatever it was, there wasn't much of any bump or anything, so the salt couldn't get hooked on much, so that's swept away easily.

Nigel: This one is easier because it has less bumps (points at fine-grain sandpaper). And this one is the hardest (points at coarse-grain sandpaper), because it has more bumps. And this one is in the middle.

Adrian's and Nigel's explanations were very similar to Mary's when she was also using a top-view drawing. All three explanations were based on the number of bumps as the central argument. On the other hand, groups that used a side-view drawing gave a different explanation. At the time when the lesson occurred, neither the students nor the teachers noted that the two explanations had very different meanings.

At this point, our analysis on multiple representations has identified two sequences of re-representation, which led to two different explanations. It has also led us to identify the group poster (i.e., Figure 3) as a representation that required further in-depth analysis. Although this initial analysis gave us an overview over a broad timescale, it was not able to tell us *how* the two sequences led to the production of different meanings by the students. According to our framework (see lower portion of Figure 1), a fine-grained analysis of the semiotic affordances of both the top- and side-view drawings is required to understand how the two different explanations for the phenomenon of self-cleaning were generated. This is the focus of the multimodal analysis in the next phase.

Second-Phase Analysis: Semiotic Affordances

In this section, we used a multimodal approach to analyze how Mary's group made meaning with the top-view drawing, followed by that with the side view. The grain size of analysis for spoken language is at the level of a clause. Thus, the transcript in the following excerpts is divided into individual clauses. For each clause, we include a description or screen capture of their nonverbal actions (e.g., gestures, direction of gaze) that were captured in our video data. The grain size of analysis for visual representations is at the level of components (e.g., a curve representing a bump or a dot representing a grain of salt).

The following excerpt shows the interaction between the researcher (R), Mary (M), and Luke (L) after they had completed the poster. As shown from the gesture and gaze column in the excerpt, Mary was oriented to the top-view drawings on the poster as she responded to the researcher.

Excerpt 1: Analysis of meaning-making with the top-view drawing (22:56–23:45)

		Verbal Utterances	Gesture/Gaze
1	R	ok, now that you have drawn all three,	
2		can you compare what you have drawn?	
3		and can you explain why this is the salt is easier to come out,	points to fine top-view
4		for this it is harder to come out?	points to coarse top-view
5	M	oh, it's already. already said it	
6		for fine, the surface is less bumpier	reads from text written above fine top-view
7		so it's easier to scrape off	
8		the surface. The surface has less bumps than coarse	reads from text written above medium top-view
9	R	can you use your diagram to explain this? I mean	
10	M	it is there I guess	points to text above fine top-view
11	R	so can you use this to explain to me why is it easier?	points to fine top-view
12	M	oh, the surface is less bumpier	points to text above fine top-view
13		so it's easier to scrape off	
14		yet the salt has nowhere to hide	points to a red dot inside fine top-view

Using Halliday's (1994) linguistic analysis, we observed that there are two main semantic relationships in Mary's explanation. The first is an attributive relationship between an object (i.e., surface) and one of its attributes (bumpiness), as seen in "less bumpier" in (6) and (12). The second is a possessive relationship between a carrier (i.e., surface) and its possessions (i.e., bumps), as seen in "has less bumps" in (8). Next, using Kress and van Leeuwen's (2006) visual grammar, we analyzed the top-view drawings (see Figure 7). Each drawing realizes a possessive relationship between a carrier (rectangular boxes) and its possessions (curves and dots). In a linguistic sense, each drawing is saying that there are bumps and salt (represented by curves and dots) inside the sandpaper. Furthermore, we saw that the students drew progressively more curves and dots for the fine-, medium-,

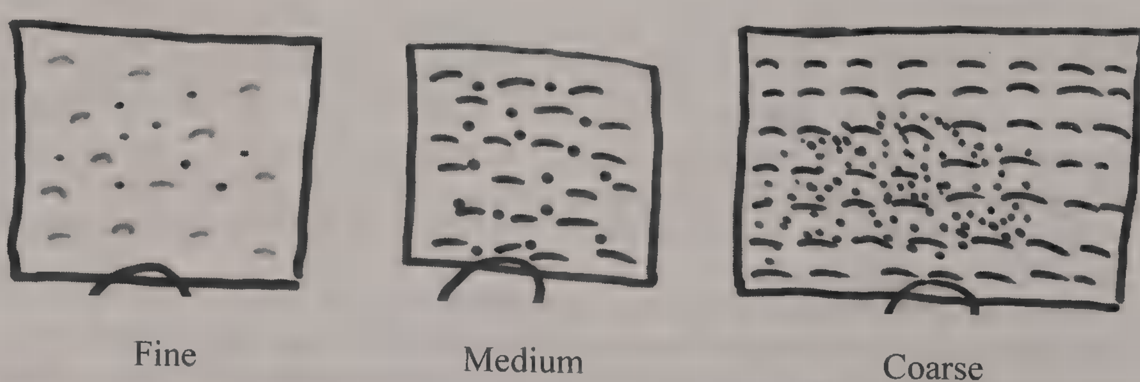


Figure 7. Top-view images of sandpapers with bumps and salt.

and coarse-grain sandpaper. Both the linguistic and visual analyses indicate the possessive relationship.

Complementing the visual analysis with the linguistic analysis, we can infer that the word "bump" means a protruding peak that the students must have imagined the sandpaper to have, based on and re-represented from their sensory experiences with the sandpaper. We also infer that students understand "bumpiness" as a quality that arises from the *number* of protruding peaks. Students' reasoning when using the top-view representation is that the number of peaks (bumps) determines the ease with which the bacteria can be scraped off a surface. We call this the *argument of quantity*.

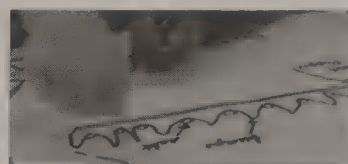
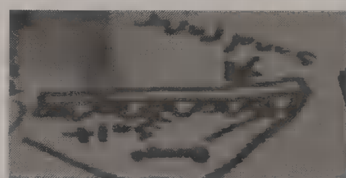
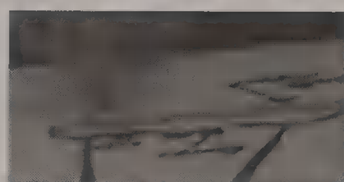
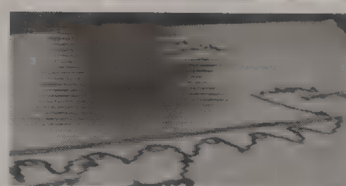
In the next analysis, we analyze and compare how the students' argument changed as they used the side-view drawing. The following excerpt occurred shortly after Excerpt 1. A crucial turn of events here was that the researcher started pointing at the side-view drawings on the poster (18–20). Consequently, Mary (M), Luke (L), and Dave (D) were oriented to the side-view drawings on the poster as they responded to the researcher (R).

Excerpt 2: Analysis of meaning-making with the side-view drawing (24:18–25:05)

		Verbal Utterances	Gesture/Gaze	Video Snapshots
18	R	So what's the difference between this bump	points to medium side view	
19		this bump	points to coarse side view	
20		and this bump?	points to fine side view	
21	M	this one you can get into	points to fine side view	
22		this one you can't get into	points to coarse side view	
23		and this [one	points to medium side view	
24	L	[you could		
25	D	[you sort of can		
26	R	but you say you can't get into, cannot get into what?		
27	M	get into [the salt]	points to medium side view	
28	L	[the salt gets more	hand gestures downwards over medium side view	
29	M	gets to [the salt and		
30	L	you can get into the salt. the salt can gets into the sandpaper		
31		but it can get into this one	fingers land on coarse side view	
32	R	But if you say the salt cannot get in, but this is a salt right?	points to a circle in coarse side view	

Continued

Excerpt 2: Continued

		Verbal Utterances	Gesture/Gaze	Video Snapshots
33		It's inside		
34	M	yah. because when you're scraping,	points to peaks of coarse side view	
35		you're scraping the top of the bumps		
36		and you can't get into it	gestures scraping motion	
37		this one you can get a little bit into the salt	points to a groove in medium side view	
38		And this one [you can get it]	points to a groove in fine side view	
39	L	[this] is the sandpaper right here	traces the length of coarse side view	
40		it's trying to get down here	traces downward motion into a groove of coarse side view	
41		So basically it's not all the way down		
42	M	you can't get it at all		

[] indicates start and end points of overlapping speech.

Again, we began with a linguistic analysis of the key clauses. Unlike the first excerpt, there was a notable shift in the grammatical subjects in clauses (21–22), (30–31), (36–38), and (42). Instead of "surface" (e.g., *surface* has less . . .), the main grammatical subjects in the second excerpt are "you" or "it," which refers to the scraper (e.g., *you* can/can't get into). This corresponds to a shift from the earlier possessive relationship of the sandpaper (e.g., surface *has* bumps) to a different kind of semantic meaning. This meaning focuses on the transitive action (Halliday, 1994) of the scraper *doing* something *to* the salt/sandpaper (e.g., you can *get into* it).

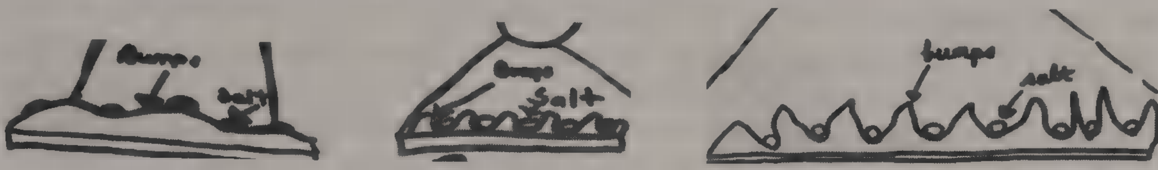


Figure 8. Side-view images of sandpapers (carrier) with bumps (possessive attributes).

A visual analysis shows that the side-view drawings (see Figure 8) retain the possessive relationship between a surface carrier and its possessions, as in the first excerpt. Thus, one of the key differences in the students' explanation in this excerpt is that each of the modalities plays a different and complementary role: The verbal modality realizes the action of the scraper while the visual realizes the surface's possessive attributes of grooves of varying depths. As for the gestural modality, there are also some differences compared to the first excerpt. First, the students were pointing (deictic gestures) more specifically at a smaller part of the sandpaper drawings (see clauses 35, 37, and 38). Second, the students used some kind of actions to animate by resemblance (iconic gestures) the physical movement of the scraper in removing the salt (see clauses 36 and 41). On the other hand, few gestures were used in the first excerpt.

Collectively, the gestural, verbal, and visual modalities resulted in a different meaning. From the repeated and synchronized uttering of the phrases "get into," "get it," or "get down," with the deictic and iconic gestures (see 35–41) referencing the side-view drawings, we can infer that the students' argument centers on the varying depth of the sandpaper "bumps." We call this the *argument of depth*. From the point of view of the curriculum, this argument is a more accurate form of reasoning for the concept of self-cleaning as compared to the earlier argument of quantity. In fact, the fine sandpaper has many more bumps than the medium or coarse sandpaper, so the argument of quantity that states that the fine sandpaper "has fewer bumps" misrepresents the phenomenon being modeled.

While the large grain-sized re-representation analysis showed that the two different representations led to different explanations, this fine-grained semiotic affordances analysis revealed *how* the representations supported the different explanations. Although both the top- and side-views realize a similar possessive relationship of a carrier (sandpaper surface) and its possessions (bumps), there is one crucial difference between the two representations. The top view is what Kress and van Leeuwen (1996) call an *inclusive* analytical structure, which shows only *some* of the possessive attributes as bumps and the rest of the surface of the carrier as blank space. On the other hand, the side view is an *exclusive* analytical structure that shows the *entire* surface of the carrier covered by the possessive attributes. Critically, in a side-view representation one cannot draw bumps without also drawing grooves on the surface. By contrast, in a top-view representation, bumps can be drawn without the grooves on the surface. By choosing to represent through a side view, both the protruding bumps and depressed grooves will be included in the representation. Therefore, although both the top- and side-view representations may appear to refer to the same phenomenon, each has different semiotic affordances that allow and constrain different kinds of meanings and argument that can be made in conjunction with the contextualizing utterances and gestures.

While research on multiple representations recognizes that different representations may constrain, complement, or help construct meaning (Ainsworth, 2006), the analysis that we presented here develops and explains *how* this meaning-making process occurs. Our analysis reveals how the processes at a shorter timescale (e.g., dialogue and gestures around the side-view representation) build up to the processes at a longer timescale (creation of

the group poster and subsequent presentation of the group's ideas). We thus show how the fine-grain, short timescale analysis based on the multimodal representations literature can inform the large-grain, long timescale analysis based on the multiple representations literature. Analyzing the inclusive/exclusive analytical structure of these representations—from a multimodal perspective—helps explain *how* the side-view representation captures important features of the phenomenon that the top-view drawing does not, thus greatly enriching the understanding of the sequence of re-representation. On the other hand, a fine-grained multimodal analysis of the group poster alone would miss important contextual information from the large-grained re-representation analysis. For instance, the physical experiment explains why the students talked about "bumps," the individual sketches explain why the group poster initially included the top-view representations, and the diagram drawn by the teacher on the whiteboard explains why the students later drew a side view. It is precisely these connections across temporal and compositional dimensions that our integrative framework allows.

Analysis on the Sixth Lesson

In this section, we move ahead to the last day of the lesson unit to show another example of our analysis. As the analytical process is very similar to what has been illustrated earlier, we will provide only the main results from our analysis.

On the sixth and final day of the lesson unit, each group of students was working to prepare a 3-minute skit to advertise the product of a nanotoilet—a recent commercial application of self-cleaning in a bathroom accessory. During the brainstorming phase, which occurred on the fifth lesson, students decided within their groups what representations they would use in their skit. Mary's group decided to create a poster. From the analysis on the multiple representations used, the sequence of re-representations over the 2 days was a commercial advertisement of a toilet that used a supersmooth nanotech finish, a written rubric, a group poster, and a video of their skit. As shown in Figure 9, their poster included side views comparing a conventional toilet to the smoother nanotoilet. They did not use the top-view representations. This shows that they recognized the top view was not useful for their explanations.

We selected the group poster and its corresponding video segment for a fine-grained multimodal analysis so that a comparison could be made with the earlier multimodal analysis of the poster from the first lesson. The linguistic analysis of the students' explanation

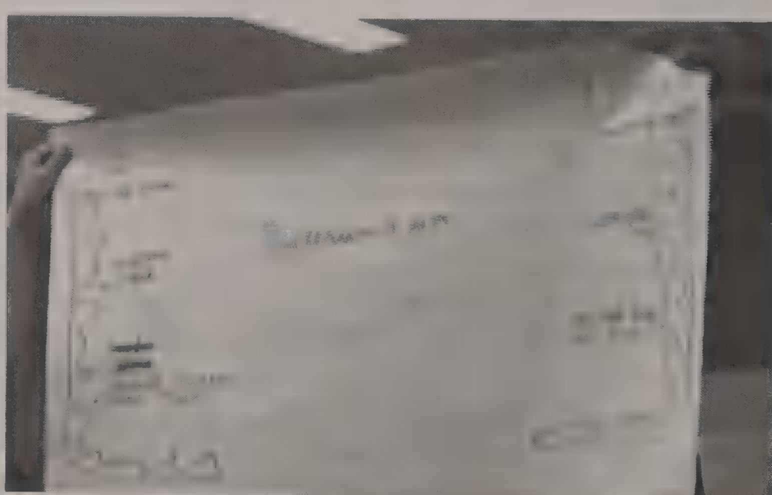


Figure 9. Group poster created by Mary's group.

(which was collaboratively made by Mary and Luke) revealed two semantic relationships that were also made during the first lesson (see Excerpt 2). As seen from the following utterances, one was the possessive relationship of a carrier and its possessive attributes ("toilet has invisible bumps"), and the other a transitivity action of the agent removing the residual bacteria ("water flows," "bacteria to come straight off"):

> Mary: The new *toilet has invisible bumps* too, but they are thirty nanometers, and nanometer is thousand times smaller than a micrometer. So a bacterium won't fit in that bump that small.
>
> Luke: It would be smoother for the *bacteria to come straight off* when the *water flows* when you flush the toilet. (emphasis added)

Again, the verbal, visual, and gestural semiotic modalities play different and complementary roles in the overall construction of their explanation, which reiterated their argument of depth from the first lesson. The group also incorporated several additional concepts they had picked up during intervening lessons. One important concept that the students learned on the third lesson was the mathematical relationship between a millimeter, micrometer, and nanometer; other information was the size of a bacterium (see Figure 9) and of the surface features of the nanotech toilet (e.g., "invisible bumps"; see excerpt above). Mary was then able to provide convincing evidence for her assertion that a bacterium "won't fit" into the surface features of the new nanotoilet. Through this mathematical reasoning, she was able to construct what we call the argument of relative size.

DISCUSSION

Comparing the analyses of the two phases with different timescale and grain size, we found that each analysis plays a mutually complementary role in illuminating students' learning with representations. In the analysis of multiple representations with a long timescale and large grain size, we learned how the different sequences of re-representations led to the production of two different representations of surface features (top vs. side view). We showed that several groups produced two different explanations depending on the representation they focused on. We also showed how the re-representation process incorporated the interaction and context of preceding activities and projected them into subsequent activities. However, this analysis did not shed light on *how* the students made meaning with the representations, nor *why* the explanations differed. In the fine-grained analysis of multimodal representations, we learned how the students used the top- and side-view drawings, along with their utterances and gestures, to construct different meanings. Although both representations portray bacteria being trapped within the cracks or pits of different surfaces, their semiotic affordances are different and supported different explanations. Thus, to understand how meanings emerged through the situated use of representations, a fine-grained analysis of the composite parts of a representation and how they were integrated multimodally by the learners was undertaken. At the same time, if the analysis was carried out only on a single representation, we would miss important details in the overall understanding of the learning process in this lesson unit. Thus, the multiple representations analysis complements the fine-grained analysis by providing this contextual information. The complementary roles between multiple and multimodal representations is summarized in Figure 10.

We propose that our integrative framework is a step toward the goal of a unified multirepresentational framework envisioned by Yore and Treagust (2006). We have shown how our framework integrates the research on multiple representations (archetypically in

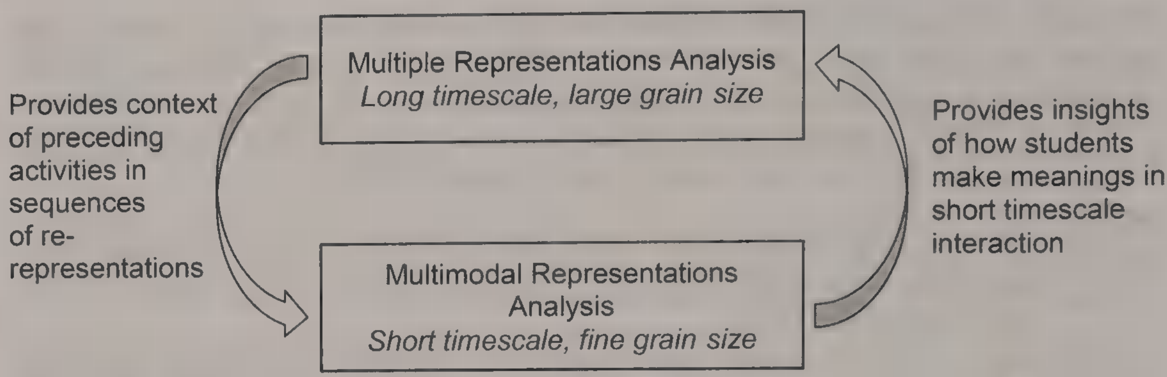


Figure 10. Complementary relationship between multiple and multimodal representations analysis.

the top left quadrant of Figure 1) and research on multimodal representations (archetypically in bottom right quadrant of Figure 1). We next show how our integrative framework suggests promising new directions for the analysis of studies that fall into the other two quadrants of Figure 1. For instance, we note that many of the studies with large grain size but short timescale (upper right quadrant) are multiple representation analyses that focus on the relative effectiveness of single representations rather than longer instructional episodes employing many representations. These studies can benefit from a multimodal analysis to begin determining *why* a given configuration or type of representation is more effective than another. For instance, the contiguity principle states that narration in words and images should be simultaneous rather than sequential so that it is easier for a learner to build connections in his/her working memory (Mayer, 2001). A fine-grained multimodal analysis can elaborate the processes by which those connections are made, as we showed above for the case of self-cleaning surfaces. Studies with small grain size but longer timescales (bottom left quadrant) are multimodal analyses focusing on multiple teaching/learning episodes across time. Such studies can add a layer of analysis that examines the sequential re-representation process; analyzing how students transform one representation into another in situated social activities, in addition to analyzing how the composite elements within and across representations interact to support the meaning-making process.

CONCLUSION AND IMPLICATIONS

Our integrative framework can inform future research on science learning with and through representations. From Vygotsky's (1986) sociocognitive theory, researchers have come to broadly understand representations as symbolic tools that mediate social learning and human cognition (e.g., Bransford, 2000; Kozma, Chin, Russell, & Marx, 2000). However, the mechanism by which this occurs is still not well understood. Our framework suggests the importance of considering re-representation as well as semiotic affordances in the analysis of students' learning with representations. This implication for research has a parallel implication for practice: For students to develop better scientific understanding, they must engage more actively in the construction of representations (Hubber et al., 2010; Waldrip, Prain, & Carolan, 2010), as they did in the summer program we studied.

In this paper, we have shown how our integrative framework can be used *retrospectively* to analyze student representation practices and artifacts. Future work is needed to explore how our framework can be used *prospectively* in the design of curriculum and instruction. Our inclusion of side-view and top-view representations in the curriculum materials was based on general multiple representations principles, and the research value of a fine-grained

examination of the multimodal representations emerged rather than being purposefully designed into the materials. Having observed that some groups did not generate the argument of depth and relative size as a result of not using a side-view way of representing, we consequently realized the importance of *building in* opportunities for students to engage more deeply with the multimodal components of representations. We expect that our framework and supporting case study will provide guidance in developing future materials that can better support student learning, in addition to future research leading to the iterative refinement of this unified multirepresentational framework.

We wish to express our gratitude to the students, teachers, and administrators of the school district who collaborated in this study. We also acknowledge and thank Jay Lemke, Joseph Krajcik, Chris Quintana, and Leema Berland for their comments and suggestions on the analyses presented in this paper. Any opinions expressed in this work are those of the authors and do not necessarily represent those of the funding agencies.

REFERENCES

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16(3), 183–198.

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. Journal of the Learning Sciences, 11(1), 25–61.

Airey, J., & Linder, C. (2009). A disciplinary discourse perspective on university science learning: Achieving fluency in a critical constellation of modes. Journal of Research in Science Teaching, 46, 27–49

Baldry, A., & Thibault, P. J. (2006). Multimodal transcription and text analysis. London: Equinox.

Bransford, J. D. (2000). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

Chandler, P., & Sweller, J. (1996), Cognitive load while learning to use a computer program. Applied Cognitive Psychology, 10, 151–17.

Fredlund, T., Airey, J., & Linder, C. (2012). Exploring the role of physics representations: An illustrative example from students sharing knowledge about refraction. European Journal of Physics, 33, 657–666.

Gilbert, J. K., & Treagust, D. F. (Eds.). (2009). Multiple representations in chemical education. Dordrecht, The Netherlands: Springer.

Halliday, M. A. K. (1978). Language as social semiotic: The social interpretation of language and meaning. London: Edward Arnold.

Halliday, M. A. K. (1994). An introduction to functional grammar. London: Arnold.

Hand, B., Alvermann, D. E., Gee, J., Guzzetti, B. J., Norris, S. P., Phillips, L. M., et al. (2003). Message from the "Island group": What is literacy in science literacy? Journal of Research in Science Teaching, 40, 607–615.

Hand, B., Gunel, M., & Ulu, C. (2009). Sequencing embedded multimodal representations in a writing to learn approach to the teaching of electricity. Journal of Research in Science Teaching, 46(3), 225–247.

Hubber, P., Tytler, R., & Haslam, F. (2010). Teaching and learning about force with a representational focus: Pedagogy and teacher change. Research in Science Education, 40, 5–28.

Jaipal, K. (2009). Meaning making through multiple modalities in a biology classroom: A multimodal semiotics discourse analysis. Science Education, 94, 48–72.

Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. Learning and Instruction, 13, 205–226

Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. Journal of the Learning Sciences, 9(2), 105–143.

Krajcik, J., & Blumenfeld, P. C. (2006). Project-based learning. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 317–334). New York: Cambridge University Press.

Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2001). Multimodal teaching and learning: The rhetorics of the science classroom. London: Continuum.

Kress, G., & van Leeuwen, T. (2006). Reading images: The grammar of visual design (2nd ed.). London: Routledge.

Lemke, J. L. (1990). Talking science: Language, learning and values. Norwood, NJ: Ablex.

Lemke, J. L. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. Martin & R. Veel (Eds.), Reading science (pp. 87–113). London: Routledge.

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. Mind, Culture and Activity, 7(4), 273–290.
Márquez, C., Izquierdo, M., & Espinet, M. (2006). Multimodal science teachers' discourse in modeling the water cycle. Science Education, 90, 202–226.
Mayer, R. E. (2001). Multimedia learning. Cambridge, England: Cambridge University Press.
McNeill, D. (2005). Gesture and thought. London: University of Chicago Press.
Pozzer-Ardenghi, L., & Roth, W. M. (2007). On performing concepts during science lectures. Science Education, 91, 96–114.
Prain, V., Tytler, R., & Peterson, S. (2009). Multiple representation in learning about evaporation. International Journal of Science Education, 31(6), 787–808.
Prain, V., & Waldrip, B. (2006). An exploratory study of teachers' and students' use of multi-modal representations of concepts in primary science. International Journal of Science Education, 28, 1843–1866.
Rosengrant, D., Etkina, E., & Van Heuvelen, A. (2007). An overview of recent research on multiple representations. AIP Conference Proceedings, 883(1), 149–152.
Roth, W.-M., & Tobin, K. (1997). Cascades of inscriptions and the re-presentation of nature: How numbers, tables, graphs, and money come to re-present a rolling ball. International Journal of Science Education, 19(9), 1075–1091.
Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. Learning and Instruction, 13, 141–156.
Tang, K.-S., & Moje, E. (2010). Relating multimodal representations to the literacies of science. Research in Science Education, 40, 81–85.
Tang, K.-S., Tan, S. C., & Yeo, J. (2011). Students' multimodal construction of work-energy concepts. International Journal of Science Education, 33, 1775–1804.
Vygotsky, L. (1986). Thought and language. Cambridge, MA: MIT Press.
Waldrip, B., Prain, V., & Carolan, J. (2010). Using multi-modal representations to improve learning in junior secondary science. Research in Science Education, 40, 65–80.
Yore, L. D. & Hand, B. (2010). Epilogue: Plotting a research agenda for multiple representations, multiple modality, and multimodal representational competency. Research in Science Education, 40, 93–101.
Yore, L. D., & Treagust, D. F. (2006). Current realities and future possibilities: Language and science literacy—empowering research and informing instruction. International Journal of Science Education, 28, 291–314.

Science
Education

SCIENCE STUDIES AND SCIENCE EDUCATION

The Distinction Between Experimental and Historical Sciences as a Framework for Improving Classroom Inquiry

RON GRAY
Center for Science Teaching and Learning, Northern Arizona University, Flagstaff, AZ 86011, USA

Received 15 July 2013; accepted 15 November 2013
DOI 10.1002/sce.21098
Published online 10 January 2014 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Inquiry experiences in secondary science classrooms are heavily weighted toward experimentation. We know, however, that many fields of science (e.g., evolutionary biology, cosmology, and paleontology), while they may utilize experiments, are not justified by experimental methodologies. With the focus on experimentation in schools, these fields of science are often not included in the inquiry experiences our students receive. I propose utilizing the distinction between experimental and historical sciences as a way to improve the diversity of scientific methodologies represented in the science classroom. This distinction can provide a framework for teachers to examine their own inquiry practices in light of the diverse methodologies present in science today. In this paper, the framework is presented and analyzed in light of the scientific practices highlighted in the Next Generation Science Standards and key concepts needed to discuss historical science methodologies are discussed. © 2014 Wiley Periodicals, Inc. *Sci Ed* **98:**327–341, 2014

INTRODUCTION

The development of classroom inquiry experiences requires an understanding about how scientists actually do their research. However, it is neither possible to depict accurately

Correspondence to: Ron Gray; e-mail: ron.gray@nau.edu
A previous version of this paper was presented at the North American Research in Science Teaching (NARST) annual conference in April 2013 in San Juan, PR.

how science works in its full complexity nor necessary. Ultimately, decisions need to be made about how to describe science in the curriculum, and it has been established that the selection of what general statements about science should be taught in schools is shaped by the social and political factors, and that what we teach as a result has consequences for how science and the public interact (Rudolph, 2003a, 2003b).

These decisions impact the way in which students come to view science as a whole and the relationships between the various fields of science. Science educators have primarily, although largely implicitly, focused on physics as the model for science as a whole. Studies in physics largely rely on experimentation[1] in which controls and variables are chosen prior to conducting the study. Replication is often required as well. Typically left out of the portrait of scientific practices are the fields of science that do not rely on experimental justification. These include historical sciences, such as geology, evolutionary biology, and cosmology, that developed methodologies to cope with problems that cannot be solved experimentally. What we often highlight as authentic science in classroom inquiry activities implies that those left out are some how less legitimate.

The relative exclusion of nonexperimental sciences has important implications. For example, evolutionary biology has been singled out by creationist groups as being unscientific on the grounds that it cannot be justified experimentally (at least at the macrolevel) but instead relies on multiple lines of observational evidence. Rudolph and Stewart point out that

> past characterizations of science, historically derived from physics, internalized a broadly empirical and experimental bias that failed to accommodate key issues evolutionary biology introduced to the scientific community. There has been little done over the past century to reconcile these views, especially in science education, resulting in a situation that continues to impede the development of effective instruction in evolution. (1998, p. 1070)

In addition, geology has historically been thought of as a derivative of physics and therefore neglected in its treatment by philosophers and historians of science and neglected as well, more importantly, by authors of the modern school curriculum (Dodick & Orion, 2003a, 2003b). This misunderstanding of the unique methodologies utilized in geology has ultimately led to limited engagement with geology for students. Owing to the integrated nature of scientific knowledge, this lack of opportunity to learn geology ultimately affects students' understandings of other subjects such as evolution, ecology, astronomy, and climate science.

At a minimum, therefore, two different modes of scientific inquiry—experimental and historical—can be distinguished and need to be considered in teaching about how scientific knowledge is constructed. While further modes are certainly discernable, the distinction of these two, originally proposed by philosophers and practitioners of science and described

[1]Diamond (1986) defines three types of experiments: laboratory experiments, field experiments, and natural experiments. While they vary in terms of their affordances and drawbacks (e.g., regulation of independent variables, generalizability, etc.) they are all attempts at singling out specific variables to examine their effect. They are all three tools that are used across a broad array of scientific disciplines. In this paper, experimentation is defined broadly as the manipulation of nature to test the relationships between variables. It is not uncommon for the term experiment to be used more broadly to refer to any type of investigation (whether a true experiment or an observational study) or, in some cases, any type of simple laboratory technique. For example, in an introductory chemistry textbook (Averill & Eldredge, 2006) describing the process of discovery of the asteroid hypothesis for the Cretaceous extinction as an example of the scientific method, the authors describe "examin[ing] the ages and sizes of known impact craters in seabeds near North and South Americas" as an experiment.

more fully below, can provide a meaningful start to emphasizing the variety of scientific methodologies utilized in various science fields, which is a good first step in developing a sophisticated knowledge of epistemic process that is critical to scientific literacy. The purpose of this paper is to provide a background understanding of experimental and historical sciences and to examine the possible affordances this distinction has for providing teachers and students a more authentic view of scientific practice in its entirety. The scientific practices defined in the new U.S. standards document will be examined through this lens as will the language of school inquiry.

DEFINING TWO TYPES OF SCIENCE

Providing a reasonably authentic context for science learning requires a greater understanding of the actual methods of inquiry as practiced in diverse disciplines. Since the late 1970s, ethnographic studies have documented the activities in a variety of sciences (e.g., Latour & Woolgar, 1986; Traweek, 1992). In addition, cognitive scientists have studied the cognitive process of science demonstrated in laboratory settings (Dunbar, 1995; Klahr & Dunbar, 1988; Nersessian, 1992; Nersessian, 2009) and in scientific research artifacts (e.g., research notes and diaries; Giere, 1988; Nersessian, 1992). With few exceptions (e.g., Bowen & Roth, 2006; Latour, 1999; Roth & Bowen, 2001), the vast majority of these studies have addressed experimental sciences such as physics and chemistry while leaving out historical sciences that utilize observational data as a primary source of evidence. It is my contention that such a concentration on experimental sciences leads to the mistaken impression that science disciplines in general operate following experimental methodologies and types of reasoning employed.

Beginning with Whewell in 1837, many scientists have written on the unique epistemological and methodological challenges faced in the historical sciences. Stephen J. Gould, in particular, prolifically championed the existence of a distinct historical method in the sciences (Gould, 1986, 1989, 2002) as did others in evolutionary biology (Mayr, 1985), paleontology (Erwin, 2011), geology (Schumm, 1998), and anthropology (Diamond, 1997). Not all scientists have held a favorable view of the historical sciences, however. Ernest Rutherford famously quipped that "science is either physics or stamp collecting" (cited in Dott, 1998). Similarly, Lord Kelvin asserted, "nothing is science if it cannot be quantified" (cited in Dott, 1998). More recently, Henry Gee, a senior editor at *Nature*, stated that the assumptions we make about evolution, for example, are "baseless" (1999, p. 2). Owing to the vastness of geological time, he wrote, historical sciences are "subjective ... as they can never be tested *by experiment*, and so they are unscientific. They rely for their currency not on scientific test, but on assertion and the authority of their presentation" (p. 5, emphasis added). With respect to paleontology, Gee reminds us that

> Testability is a central feature of the activity we call science. Some have sought a kind of special dispensation for paleontology as an "historical" science, that it be admitted to the high table of science even though paleontologists cannot, classically, do the kinds of experiments other scientists take for granted. You cannot go back in time to watch the dinosaurs become extinct or fishes crawl from the slime to become amphibians ... The problem is that what we see before us is the result of a once-only experiment in history. Because it happened only once, it is not accessible to the reproducibility scientists usually require ... To see paleontology as in any way "historical" is a mistake in that it assumes that untestable stories have scientific value ... No science can ever be historical. (p. 8)

Here Gee exemplifies the traditional view of science holding experimentation and replication as the hallmarks of appropriate justification. Even physicist Luis Alvarez, codiscoverer of the asteroid hypothesis for the Cretaceous mass extinction (an exemplar of historical science), disparaged fields like paleontology for not doing "real science" (cited in Gould, 1989).

Philosophers of science have taken on these critiques against the historical sciences (Brown, 2011; Cleland, 2002; Jeffares, 2009; Kosso, 2001; Tucker, 2011; Turner, 2007).[2] Cleland (2001, 2002) first argued against the claim that historical sciences are methodologically inferior on the grounds of the asymmetry of overdetermination. She claims that historical sciences have evolved methodologies to cope with the overdetermination of past events (a single event in the past leaves multiple traces of evidence with which we can infer the original event), whereas experimental sciences have similarly evolved other methodologies to cope with the underdetermination of future events (multiple causes are possible for a single event in the present). Therefore, as each "selectively exploits the differing information that nature puts at its disposal, there are no grounds for claiming that the hypotheses of one are more securely established by evidence than are those of the other" (2001, p. 990). While the details have been and are still actively debated among philosophers of science,[3] the distinction between experimental and historical sciences has been a productive one.

Although the discussion over what truly separates these two modes of science continues, some areas of consensus have emerged at a level of generality relevant for science education contexts. Diamond (1997) first summarized the features that set the two apart as methodology, the role of prediction, causation, and complexity.[4] Building on the previous work of Dodick, Argamon, and Chase (2009), these features have been further elucidated here as "epistemic goal," "nature of phenomena under study," "method of evidence construction," and "quality standard" (see Table 1).

Experimental sciences (e.g., chemistry, physics, molecular biology) ask questions in which direct experimentation of natural phenomena is possible. Therefore, in these sciences, knowledge is most often constructed through controlled experiments in which natural phenomena are manipulated, often to test a single hypothesis[5] (*method of evidence construction*) (see Table 1). For example, Ernest Rutherford's Geiger–Marsden experiment in which positively charged alpha particles were directed at a thin layer of gold foil tested the prevailing "plum pudding" model of the atom by revealing the existence of the atomic nucleus. Hypotheses are evaluated on the consistency between predictions and experimental results, reproducibility, and generalizability to a wide range of phenomena in multiple contexts (*quality standard*). The goal of these sciences is to find general laws or statements (e.g., kinetic molecular theory) (*epistemic goal*) that are made possible by the uniformity of the objects under study (e.g., atoms) (*nature of phenomenon under study*). In other words, these sciences concern natural events that are general and repeated easily. Rarely do the

[2]While the terms experimental and historical sciences are often used in the philosophy of science, other terms have included inductive vs. deductive, nomothetic vs. historical, analytical vs. synthetic, and demonstrative vs. nondemonstrative (Rudolph & Stewart, 1998; Sober, 1993).

[3]For example, Turner (2007) doubts Cleland's claim about the asymmetry of overdetermination. Instead, he argues that the distinction between the two types of sciences lies in the asymmetry of background theories (among others). More recently, he has also questioned Cleland's definition of prediction and called attention to the exclusion of scientific practices such as inference from pattern to process and modeling in her analysis (2013).

[4]See Diamond's popular book *Guns, Germs, and Steel: The Fates of Human Societies* (1997, pp. 420–425) for a noteworthy discussion of the historical sciences.

[5]In addition to hypothesis testing, experiments can also be conducted to measure a parameter (e.g., Millikan's oil drop experiment to measure the charge of an electron) or simply to describe an aspect of nature.

TABLE 1
Comparing Experimental and Historical Sciences (Adapted From Dodick et al., 2009)

	Experimental Sciences	Historical Sciences
Epistemic goal	To find general laws or statements of natural phenomena (e.g., kinetic molecular theory)	To find causes of past phenomena from present traces of evidence (e.g., plate tectonics as a cause of various geological phenomena)
Nature of phenomena under study	Uniform and interchangeable entities (e.g., atoms)	Complex and unique entities (e.g., the big bang)
Method of evidence construction	Manipulation of natural phenomena to test a single (often complex) hypothesis (e.g., the Geiger–Marsden experiment)	Observation of natural phenomena (often to test multiple competing hypotheses) (e.g., measurement of iridium in the K–T boundary)
Quality standard	Effective prediction (e.g., prediction of the degree to which light bends around the sun to test relativity theory)	Effective explanation (e.g., the large variety of facts explained by evolutionary theory)

particularities of place and time play a significant role in the reasoning process (Frodeman, 1995).

The historical sciences (e.g., paleontology, cosmology,[6] evolutionary biology), on the other hand, gather evidence by observation because direct experimentation is usually impossible (*method of evidence construction*). These sciences most often utilize observational evidence, what Cleland (2002) refers to as evidentiary "traces from the past," to investigate ultimate causes from the past whose effects must be interpreted from complex, causal chains of events (*epistemic goal*) (Mayr, 1985). For example, Alfred Wegener used multiple pieces of evidence (biogeography of extinct organisms, the complementary arrangement of continents, patterns in glacial sedimentation, etc.) to argue for the theory of continental drift. Thus, the quality of this research is often based on the adequacy of the explanation (*quality standard*) rather than successful prediction[7] since it is based on the study of complex and unique entities (e.g., the big bang) that have a low probability of repeating exactly (if at all) (*nature of phenomenon under study*). One can rarely be assured that any two examples of a past phenomenon are exactly the same. In other words, these sciences attempt to construct causal explanations for unique events (often in the past) using multiple lines of evidence in lieu of direct experimentation. In addition, reasoning in historical sciences consists largely of explanatory or reconstructive reasoning compared to predictive reasoning from causes to effects as is found in the experimental sciences (Diamond, 1997; Gould, 1986).

It is important to note that, while knowledge claims made in the historical sciences are not justified experimentally, historical scientists do conduct experiments and utilize laboratory

[6]See Grignon (2012) for a description of cosmology as a historical science.

[7]See Cleland (2011, 2013) for a detailed comparison of predictions in experimental and historical sciences. She argues that predictions made in historical sciences are "too vague to specify precise conditions for testing and evaluating hypotheses" (p. 6).

methods. For example, experiments conducted in genetics and biochemistry have been instrumental in the development of key ideas in evolutionary biology. Radiometric-dating methods, grounded in statistical laws of quantum mechanics, are essential to theoretical understanding across a number of historical sciences. However, what are thought of as the main principles of these sciences are broad historical claims that are not open to direct testing (e.g., the big bang). In historical sciences, in other words, experiments are a means to an end as opposed to an end in itself (Jeffares, 2008).

Like all attempts at categorization, there are limitations to this interpretation of the sciences, and I do not wish to leave the distinction unproblematized. It is important to state that by no means are these two types of science to be taken as clear, distinct categories but merely as representatives of types of methods of which more could surely be determined. However, the framework proposed here does serve the larger purpose of highlighting the relative exclusion of nonexperimental methodologies in the science classroom. Within this framework, overlaps exist and debates can be had about individual fields of study as to whether or not they fit either type. For instance, Brown (2011) has recently made the case for the inclusion of ecology as a historical science. While ecology does not often involve explanation of past events, it does, according to Brown, share epistemological and explanatory characteristics with the traditional historical sciences. These shared characteristics are largely based on challenges associated with studying complex ecological systems.[8] Cleland (2002), in her influential work on the subject refers to a continuum between largely experimental disciplines (e.g., fundamental physics) and largely historical ones (e.g., paleontology). Along this continuum are disciplines such as ecology and evolutionary biology that include differing degrees of experimental and historical methodologies. Perhaps a finer grained distinction is necessary that focuses not on the disciplines or fields of study themselves but rather on the specific phenomenon under study. The two categories, as used in this paper, are meant to serve educators heuristically to highlight the large and diverse number of fields and methodologies that are commonly left out of classroom inquiry experiences. The idea is for teachers to use them to develop a more inclusive set of authentic inquiry experiences for their students.

HISTORICAL SCIENCES IN THE SCIENCE EDUCATION LITERATURE

Science education scholars recently have taken up inquiries into the historical sciences. Ault points out in his 1998 advocacy for inquiry in the geological sciences that that "geology is not physics" (p. 190) and claims that reasoning involved in explanations of geological phenomena relies on contingency and ambiguity in contrast to the generalizations that aim for prediction in experimental sciences. For example, using experimental methods chemists apply the gas laws across all contexts and are able to predict how gases react when pressure and temperature change. An explanation of earthquakes, however, requires an inference to events in the past to account for the current event. Thus, expert understanding in geology requires restricting the ambiguity inherent in inquiry about unique events. The goal is to reconstruct past geologic events and processes from observational data that cannot be recreated in a laboratory. The explanations produced through a historical mode of inquiry such as this are contingent and case dependent and are justified by the explanatory power they offer as opposed to consistency with some prediction. Similar work has been done that has revealed the features of inquiry that are distinct from experimental

[8]Kingsland (1995) provides a history of the internal struggle within the field of population ecology as scientists debated the "modernizing impulse" (p. 218) toward highly mathematical and predictive modes of reasoning.

sciences in evolutionary biology (Passmore & Stewart, 2002; Rudolph & Stewart, 1998) and paleontology (Ault & Dodick, 2010), among others.

More recently, Dodick et al. (2009) sought empirical support for the claim that the sciences do indeed rely on distinct methodological differences. They analyzed patterns of language used by scientists in 1,605 articles from 12 experimental and historical scientific journals where they found distinctive differences in language features that they were able to link to the methodological differences between the two sciences. For instance, experimental sciences use more predictive statements and binary judgment of what is possible or not. In contrast, historical sciences use a more nuanced comparison of levels of confidence in constructed explanations.

Gray and Kang (2014) analyzed the argumentation patterns of secondary science teachers during instruction on experimental and historical science units for similar rhetorical differences. Utilizing Toulmin's argumentation pattern, the authors showed that the teachers provided relatively distinct and authentic patterns of argument as would be expected based on the methodological differences. For example, the teachers included far more discussion of evidence in teaching topics based on historical inquiry than they did in teaching topics based on experimental inquiry. This would be expected as historical scientists rely on specific pieces of evidence to form a narrative explanation as their argument.

Both of these studies showed specific differences in the language used by scientists and science teachers in communicating information from the two sciences. Although science education researchers have productively utilized the distinction between experimental and historical sciences as a framework for research, little mention of this distinction, however, is presented to preservice or inservice teachers as a resource for understanding the diversity of methodologies from which to draw as they design and implement inquiry experiences in their classrooms.

NEED FOR AN EXPANDED LANGUAGE OF INQUIRY

As described by Lemke, "language is a system of resources for making meaning" (1990, p. ix). Thus the language used to describe inquiry directly influences students' understanding of it. Like the inquiry experiences students have in K-12 science classrooms, the language of the science classroom is heavily weighted toward the experimental sciences. Science teachers talk of hypotheses, predictions, experiments, controls, and variables. And while these concepts are relevant for historical sciences, they are not sufficient to represent them in a robust way. Other concepts are necessary as resources for implementing historical inquiries in the classroom. They include retrodiction, abduction, reasoning from analogy, and multiple working hypotheses (see Table 2). All of these are important in experimental sciences as well, but their exclusion from classrooms more heavily compromises student understanding of the historical sciences. The inclusion of these concepts, I argue, not only will provide better resources for examining historical sciences in the classroom but also will improve students' understanding of inquiry in the experimental sciences as well.

Retrodiction (often referred to as postdiction) is the process of inferring the past from the present (i.e., a prediction in the past). Darwin, for example, retrodicted that many intermediate forms of life would be found in the fossil record linking human beings and other primates and that similar intermediate forms would be found linking modern horses with primitive mammals. Similarly, cosmologists were able to retrodict from the big bang theory the existence of cosmic microwave background radiation. Interestingly, even though it is a process that lies at the heart of the historical sciences (Ault, 1998), the concept of retrodiction is not commonly found in the discourse about scientific inquiry within the science education community (Sibley, 2009).

TABLE 2
An Expanded Language for Classroom Inquiry

Retrodiction	• Definition: An inference about past events. • Example: Scientists tested the asteroid hypothesis by searching for impact debris, glass, shockwaves, tsunami debris, and an impact crater.
Abduction	• Definition: A type of inference in which an explanatory hypothesis is generated. • Example: Darwin presented an extended argument for natural selection as the best hypothesis for explaining the available evidence.
Reasoning from analogy	• Definition: Utilizing present causes to explain similar events in the past. • Example: The reconstruction of the locomotion and behavior of extinct animals based on similarities with extant animals.
Multiple working hypotheses	• Definition: The process by which multiple possible hypotheses are generated and systemically compared against the evidence. • Example: Hypotheses to explain the extinction of the dinosaurs included random chance, a magnetic reversal, a nearby supernovae, and volcanic activity, among others.

Whereas a retrodiction provides an inference about the past to be tested against observations, *abduction* refers to the process of forming an explanatory hypothesis (Magnani, 2003; Peirce, 1978; Walton, 2004). In contrast to induction and deduction, abduction runs backwards from effect to cause to provide a possible explanation for the manner in which the effect is observed. For example, Alfred Wegener abduced from the available evidence (e.g., fossil and geologic observations) the explanatory theory of continental drift. Thus he reasoned from the available evidence to an explanatory hypothesis from which retrodictions could be inferred to test the hypothesis.

One of the most prominent ways in which abductions occur is through the use of analogic reasoning (Frodeman, 1995). Within geology, for example, the principle of uniformitarianism implies the constancy of physical laws over time thus allowing for interpretations of past events based on current observations ("the present is the key to the past" as this is sometimes stated). Therefore current phenomena can be used as analogues to understand phenomena that occurred in the past. Paleontologists, for instance, regularly study the locomotion and other traits of extant animals (e.g., birds) to understand similar extinct animals (e.g., dinosaurs). As with all analogies, those used in reasoning about past events have limitations that must be acknowledged.

The final concept relevant in the historical sciences is that of *multiple working hypotheses* (Chamberlin, 1890). As the range of background knowledge needed for the types of reasoning presented above is so vast, and the fact that phenomena rarely result from a single cause, it is often possible that many different hypotheses can be formed at the same time. These hypotheses can then be pitted against each other as new evidence comes to light until one better explains the evidence. According to Ault (1998), the use of multiple working hypotheses produces "independent, converging lines of inquiry to evaluate the degree to which they converge upon a common solution" (p. 207). For example, prior to the seminal paper by Alvarez, Alvarez, Asaro, and Michel (1980) on the cause of the Cretaceous extinction event that eliminated an estimated 75–85% of all species on Earth, multiple hypotheses had been proposed. These included "gradual or rapid changes in oceanographic, atmospheric, or

climatic conditions due to a random or cyclical coincidence of causative factors; a magnetic reversal; a nearby supernova; and the flooding of the ocean surface by fresh water from a postulated arctic lake" (p. 1095). None of the evidence available at the time provided strong support for any one of these hypotheses over the others. During their subsequent investigation sparked by the discovery of iridium traces at the Cretaceous–Tertiary (K–T) boundary layer, Alvarez and his colleagues weighed each new piece of evidence against the multiple possible hypotheses until one, the asteroid hypothesis, was most strongly supported. In this case, the discovery of iridium in the K–T boundary rock layers was the "smoking gun."[9]

These concepts are relevant in both experimental and historical sciences. For example, in his quest for a unified treatment of the laws of black body radiation, Planck abduced the quantum hypothesis (Magnani, 2013). Similarly, while retrodiction is more common in the historical sciences, prediction does play a role as well. For example, a field geologist may predict what will be found at another location based on current observations. However, the concepts described above play a more prominent role in the historical sciences so their absence from the science classroom disproportionately affects students' understanding of the ways in which historical scientists construct knowledge in their fields.

A SHIFT TOWARD SCIENTIFIC PRACTICES

One way to effect the inclusion of these ideas into the classroom is through national standards documents that directly impact teachers' conceptions of scientific inquiry. With the recent publication of the *Framework for K-12 Science Education* (National Research Council [NRC], 2011) in the United States, the ways in which science teachers are being called upon to engage their students in authentic science has shifted. Starting in the 1960s, we moved from a focus on the methods of science to the processes of science (e.g., observing, inferring, and predicting). These processes gave us scientific inquiry as an approach to science teaching emphasizing the skills and abilities of inquiry to learn scientific concepts. These key science concepts were articulated in national science standards documents that were drafted in the late 1980s and 1990s (American Association for the Advancement of Science, 1990, 1993; NRC, 2000). Since the release of these documents, our understanding of how students learn science (Bransford & Donovan, 2005) and the way science functions (Duschl & Grandy, 2013) has progressed. Based largely on work from the science studies community, a new focus on "science as practice" has emerged which brings the doing of science and the learning of science content together. As described in *Ready, Set, Science!*, "science practice involves doing something and learning something in such a way that the doing and learning cannot really be separated" (NRC, 2000, p. 34).

The new *Framework* and subsequent Next Generation Science Standards (NGSS Lead States, 2013) highlight eight scientific and engineering practices informed by the science studies and science education literatures that are integrated with content and crosscutting concepts within the standards. The practices are (a) asking questions and defining problems, (b) developing and using models, (c) planning and carrying out investigations, (d) analyzing and interpreting data, (e) using mathematics and computational thinking, (f) constructing explanations and designing solutions, (g) engaging in argument from evidence, and (h) obtaining, evaluating, and communicating information. Taken together, they present an active view of the construction of scientific knowledge, both as it happens in science and as it should happen in the science classroom. They also represent a shift from a linear

[9]Defined by Cleland as a trace or collection of traces that "unambiguously distinguishes one hypothesis from among a set of currently available hypotheses as providing 'the best explanation' of the traces thus far observed" (2002, p. 481).

scientific methodology toward a more realistic view of the epistemic practices of scientific disciplines. They are meant to be regarded as both learning outcomes and instructional strategies.[10]

These practices are general enough to encompass inquiry across the continuum of experimental and historical sciences. In fact, the authors of the *Framework* state that they are written in a way as to not "overemphasize experimental investigation at the expense of other practices" (p. 3-2). However, in the descriptions of each of the practices the authors prioritize experimental over historical sciences. For example, in illustrating the practice of "asking questions," only questions from the experimental sciences are given as examples (e.g., how does the particle model of matter explain the incompressibility of liquids?). In fact, only examples taken from the experimental sciences are given in descriptions of the first five scientific practices (e.g., the ideal gas law, atomic theory of matter, quantum mechanics). Examples from the historical sciences first appear in descriptions of the practices of "constructing explanations" (e.g., big bang theory, theory of evolution) and "engaging in argument from evidence" (e.g., heliocentric theory, theory of evolution). Whereas the inclusion of historical science examples in these last two practices highlight the importance of major theoretical advances in all sciences, the historical sciences are entirely left out of the description of the practices that involve the design and implementation of empirical investigations once again leaving the impression that experimental methodologies are superior to historical ones in the development of scientific knowledge. Put another way, the *Framework* emphasizes through its descriptions of the practices the results of the historical sciences, but not the unique methodologies developed in these sciences to investigate phenomena. This is important because of the national standards' role in defining inquiry in science classrooms across the united States. The standards, however, do make reference to the historical sciences and provide ample opportunity for authentic inquiry experiences (e.g., "Analyze and interpret data on the distribution of fossils and rocks, continental shapes, and seafloor structures to provide evidence of the past plate motions"). Proper implementation of these standards, however, is dependent on the teachers' understanding of the scientific practices embedded within them.

All of the scientific practices are relevant to both experimental and historical sciences, and there is substantial overlap as well between the two. All sciences ask questions, analyze and interpret data, construct explanations, and so on. However, the epistemological and methodological differences between the two types of sciences (see Table 1) as well as the expanded terminology described above (see Table 2) reveal small but significant differences in the way the scientific practices may be enacted in the classroom across disciplines. All scientists, for example, ask relevant questions that define and guide their work; however, the differences in the epistemic goal and nature of the phenomena under study lead to different types of questions. The historical sciences ask questions about unique entities that cannot be manipulated experimentally (e.g., what caused the Permian extinction?), whereas experimental sciences ask questions about uniform entities for which experimentation is possible (e.g., what is the structure of DNA?). This of course affects the practice of planning and carrying out investigations as historical science investigations are largely observational and often utilize retrodictions and abductive reasoning.

In addition, the difference in the quality standard between the two sciences (effective prediction vs. effective explanation) affects the practice of analyzing and interpreting data. In experimental sciences, evidence is most often compared to a prediction, whereas in the

[10]Note that in this paper the engineering practices are not included as they are not relevant to the distinction between the two types of science.

historical sciences evidence is utilized to evaluate multiple possible hypotheses to narrow down to the most likely explanation. This also affects the explanations and arguments constructed in the two sciences. Explanations in the experimental sciences are generalizable to similar phenomena and can be used to generate further predictions. In the historical sciences, however, explanations are most often provided in narrative form[11] and are only relevant for the unique phenomenon under study. While they often include generalizations from the experimental sciences (e.g., the theory behind radiometric dating), they are not relevant as potential predictions. Arguments differ as well, mainly due to the larger amount of evidence needed to build and warrant historical arguments as compared to the limited pieces of evidence required for arguments in the experimental sciences. While certainly not an exhaustive list, these examples show that a more nuanced understanding of the practices across the disciplines is possible and, I contend, relevant to the design and implementation of historical inquiries in the K-12 science classroom.

IMPLICATIONS FOR SCIENCE EDUCATION

To present a more authentic image of scientific inquiry, science teachers need an understanding of the differences between various modes of inquiry including the experimental and historical sciences as well as images of how these differences might play out in the classroom. Some resources for teachers do already exist to aid teachers in this endeavor including inquiry-based activities derived from the historical sciences (e.g., Diamond & Zimmer, 2006; Dempsey, Bodzin, & Cirucci, 2012; Hansen & Slesnick, 2006; Kastens & Turrin, 2010; McGarry, Straffon, & Patterson, 2012).

McGarry et al. (2012), for example, constructed an activity in which students evaluate the airburst theory that posits the explosion of an extraterrestrial object in the Earth's atmosphere over North America approximately 12,900 years ago as a common cause for previously disparate phenomena (i.e., the extinction of North American megafauna, the end of the Clovis culture, the "big freeze" period of cooling, and the series of elliptical depressions called the Carolina Bays). In this activity, students evaluate the evidence for each phenomenon and construct arguments for and against the still-controversial theory. While not explicitly stated, this activity not only includes multiple scientific practices (NGSS Lead States, 2013), but retrodiction, multiple working hypotheses, abduction, and reasoning from analogy as well. With the explicit inclusion of these concepts, this activity provides a strong example of an inquiry in the historical sciences for use in the science classroom. In addition to these activities, rich descriptions of discoveries in the historical sciences are available as resources from which educators can design similar historical inquiries (e.g., Allchin, Singer, & Hagen, 1996; Atwater, 2005; Raup, 1999; Smoot & Davidson, 2007; Weiner, 1995). Models of inquiry specific to the earth sciences have been proposed (Oh, 2008; Pyle, 2008) to provide guidance for teachers and curriculum developers as well.

Even with the examples highlighted above, authentic images of inquiry in the historical sciences for use in classrooms are still the exception rather than the rule. Science teachers who teach historical science topics in the classroom must familiarize themselves with the unique methodologies of the historical sciences as well as the additional concepts and terminology historical science inquiries require. Science teacher educators are then challenged to provide authentic images of practice in the historical sciences as well as examples of

[11] Common to historical sciences, narrative explanations "construct a story—a coherent, intuitively continuous, causal sequence of events centering on a precipitating event and culminating in the phenomena in need of explanation" (Cleland, 2011, p. 17).

effective classroom inquiries. This means science methods courses must provide experience not only with implementing experiments in the classroom but implementing historical inquiries as well while highlighting the terminology needed to adequately represent the methods and reasoning inherent in these inquiries. Finally, curriculum developers are challenged to focus not only on the discoveries of the historical sciences, but also on the practices by which the discoveries were made when producing classroom resources.

CONCLUSIONS

The highly influential book *Inquiry and the National Science Education Standards* (INSES) (NRC, 2000) sought to illustrate the inquiry strand of the U.S. national standards by providing a richly described example of scientific inquiry. The authors chose as an example a classic study in the historical sciences, Nelson et al.'s (1995) discovery of a past subduction zone earthquake along the U.S. Pacific Northwest coastline. Not only did the authors provide as the only example of scientific inquiry one from a historical science, but they describe scientific inquiry as "the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work" (p. 1), a definition inclusive of the historical sciences. Even as INSES helped formalize our conception of inquiry in the science classroom, the role of historical sciences has still been largely left out of the conversation for over a decade. The new U.S. national standards provide a similarly broad and inclusive definition of scientific inquiry, yet the history of INSES shows that this is insufficient to enact change in our teachers' conceptions of authentic scientific inquiry or their abilities to enact them in the classroom.

I propose that the distinction between experimental and historical sciences provides a framework from which to more fully integrate the ways in which researchers in the historical sciences construct new knowledge. It is not enough, as is displayed in the *Framework* (NRC, 2011) and as is far more common in classrooms and curricula, to merely focus on the end products of the historical sciences. The sweeping explanatory theories of the historical sciences are certainly important, but so are the ways in which the community of scientists made those discoveries. The distinct methodologies and patterns of reasoning and arguing employed in the historical sciences needs to be included in our classroom inquiry experiences so that students can develop a richer and more complete image of science.

As described earlier, this distinction is not without its problems. There is much overlap between the experimental and historical sciences; they have far more points in common than they do points of difference. However those small differences are important since misunderstandings about the justification of historical science claims can have significant consequences as in the case of creationist critiques of evolution (Rudolph & Stewart, 1998) and the impact on geology in the U.S. K-12 curriculum (Dodick & Orion, 2003a, 2003b). As illustrated here, teachers will need an understanding of how these differences play out in terms of the shift toward "science as practice" as codified in the new U.S. national standards. They will also require an expanded vocabulary of concepts that better illustrate the methodologies of the historical sciences such as abduction, retrodiction, etc. Taken together, I believe the concepts and tools presented here provide a solid starting point from which to move, as Rudolph states, from a "facile stereotype of some non-existent, singular scientific method" to a more authentic understanding of "the process of knowledge construction as it's actually practiced" (2007, p. 3).

REFERENCES

Allchin, D., Singer, F., & Hagen, J. B. (1996). Doing biology. New York: HarperCollins College.

Alvarez, L., Alvarez, W., Asaro, F., & Michel, H. (1980). Extraterrestrial cause for the Cretaceous-Tertiary extinction. Science, 208(4448), 1095–1108.

American Association for the Advancement of Science. (1990). Science for all Americans. New York: Oxford University Press.

American Association for the Advancement of Science. (1993). Benchmarks for science literacy (p. 418). New York: Oxford University Press.

Atwater, B. (2005). The orphan tsunami of 1700: Japanese clues to a parent earthquake in North America. Seattle, WA: U.S. Geological Survey and the University of Washington Press.

Ault, C. R. (1998). Criteria of excellence for geological inquiry: The necessity of ambiguity. Journal of Research in Science Teaching, 35(2), 189–212.

Ault, C. R., & Dodick, J. (2010). Tracking the footprints puzzle: The problematic persistence of science-as-process in teaching the nature and culture of science. Science Education, 94(6), 1092–1122.

Averill, B., & Eldredge, P. (2006). Chemistry: Principles, patterns, and applications. San Francisco, CA: Pearson Benjamin Cummings.

Bowen, G. M., & Roth, W.-M. (2006). The practice of field ecology: Insights for science education. Research in Science Education, 37(2), 171–187.

Bransford, J. D., & Donovan, M. S. (2005). Scientific inquiry and how people learn. In M. S. Donovan & J. D. Bransford (Eds.), How students learn: Science in the classroom (pp. 397–419). Washington, DC: National Academies Press.

Brown, B. (2011). Ecology as historical science. Philosophy of Ecology, 11, 251–284.

Chamberlin, T. (1890). The method of multiple working hypotheses. Science, 15(366), 92–96.

Cleland, C. E. (2001). Historical science, experimental science, and the scientific method. Geology, 29(11), 987.

Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. Philosophy of Science, 69(3), 447–451.

Cleland, C. E. (2011). Prediction and explanation in historical natural science. British Journal for the Philosophy of Science, 62(3), 551–582.

Cleland, C. E. (2013). Common cause explanation and the search for a smoking gun. Geological Society of America Special Papers, 502, 1–9.

Dempsey, C., Bodzin, A., & Cirucci, L. (2012). Reconstructing environmental change using lake varves as a climate proxy. Science Scope, 35(7), 42–47.

Diamond, J. M. (1986). Overview: Laboratory experiments, field experiments, and natural experiments. In J. M. Diamond & T. Chase (Eds.), Community ecology (pp. 3–22). New York: HarperCollins College.

Diamond, J. M. (1997). Guns, germs, and steel: The fates of human societies. New York: WW Norton.

Diamond, J., & Zimmer, C. (2006). Virus and the whale: Exploring evolution in creatures small and large. Arlington, VA: NSTA Press.

Dodick, J., Argamon, S., & Chase, P. (2009). Understanding scientific methodology in the historical and experimental sciences via language analysis. Science & Education, 18(8), 985–1004.

Dodick, J., & Orion, N. (2003a). Introducing evolution to non-biology majors via the fossil record: A case study from the Israeli high school system. The American Biology Teacher, 65(3), 185–190.

Dodick, J., & Orion, N. (2003b). Geology as an historical science: Its Perception within science and the education system. Science & Education, 12, 197–211.

Dott, R. H. (1998). What is unique about geological reasoning? GSA Today, 8(10), 15–18.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. E. Davidson (Eds.), The nature of insight (pp. 365–396). Cambridge, MA: MIT Press.

Duschl, R. A., & Grandy, R. (2013). Two views about explicitly teaching nature of science. Science & Education, 22(9), 2109–2139.

Erwin, D. H. (2011). A paleontological look at history. Cliodynamics, 2(1), 27–39.

Frodeman, R. (1995). Geological reasoning: Geology as an interpretive and historical science. Geological Society of America Bulletin, 107(8), 960–968.

Gee, H. (1999). In Search of deep time: Beyond the fossil record to a new history of life (pp. 267). New York: Cornell University Press.

Giere, R. N. (1988). Explaining science: A cognitive approach. Chicago: University Of Chicago Press.

Gould, S. J. (1986). Evolution and the triumph of homology, or why history matters. American Scientist, 74, 60–69.

Gould, S. J. (1989). Wonderful life: The burgess shale and the nature of life. New York: Norton.

Gould, S. J. (2002). The structure of evolutionary theory (pp. 1433). Cambridge, MA: Harvard University Press.

Gray, R. E., & Kang, N.-H. (2014). The structure of scientific arguments by secondary science teachers: Comparison of experimental and historical science topics. International Journal of Science Education, 36(1), 46–65.
Grignon, C. (2012). Is cosmology a historical science? AIP Conference Proceedings, 1446, 333–337.
Hansen, T., & Slesnick, I. (2006). Adventures in paleontology: 36 classroom fossil activities. Arlington, VA: NSTA Press.
Jeffares, B. (2008). Testing times: Regularities in the historical sciences. Studies in History and Philosophy of Biological and Biomedical Sciences, 39(4), 469–475.
Jeffares, B. (2009). Guessing the future of the past. Biology & Philosophy, 25(1), 125–142.
Kastens, K., & Turrin, M. (2010). Earth science puzzles: Making meaning from data. Arlington, VA: NSTA Press.
Kingsland, S. (1995). Modeling nature: Episodes in the history of population ecology (2nd. ed.). Chicago: University of Chicago Press.
Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. Cognitive Science, 12(1), 1–48.
Kosso, P. (2001). Knowing the past: Philosophical issues of history and archaeology. Amherst, NY: Humanity Books.
Latour, B. (1999). Pandora's hope: Essays on the reality of science studies. Cambridge, MA: Harvard University Press.
Latour, B., & Woolgar, S. (1986). Laboratory life. Princeton, NJ: Princeton University Press.
Lemke, J. L. (1990). Talking science: Language, learning, and values. Westport, CT: Greenwood Publishing.
Magnani, L. (2003). Abduction, reason, and science: Processes of discovery and explanation. New York: Kluwer Academic Publishers.
Magnani, L. (2013). Is abduction ignorance-preserving? Conventions, models and fictions in science. Logic Journal of IGPL, 21(6), 882–914.
Mayr, E. (1985). How biology differs from the physical sciences. Cambridge, MA: MIT Press.
McGarry, M., Straffon, D., & Patterson, C. (2012). Explaining four earth science enigmas with a new hypothesis. Science Scope, 35(7), 35–41.
National Research Council. (2000). Inquiry and the National Science Education Standards: A guide for teaching and learning. Washington, DC: National Academy Press.
National Research Council. (2011). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
Nelson, A., Atwater, B., Bobrowsky, P., Bradley, L., Clague, J. J., Carver, G. A., et al. (1995). Radiocarbon evidence for extensive plate-boundary rupture about 300 years ago at the Cascadia subduction zone. Nature (London), 378, 371–374.
Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. Cognitive Models of Science, 15, 3–44.
Nersessian, N. J. (2009). How do engineering scientists think? Model-based simulation in biomedical engineering research laboratories. Topics in Cognitive Science, 1(4), 730–757.
NGSS Lead States. (2013). Next generation science standards: For states, by states. Washington, DC: The National Academies Press.
Oh, P. S. (2008). Adopting the abductive inquiry model (AIM) into undergraduate earth science laboratories. In I. V. Eriksson (Ed.), Science education in the 21st century (pp. 263–277). New York: Nova Science.
Passmore, C. M., & Stewart, J. H. (2002). A modeling approach to teaching evolutionary biology in high schools. Journal of Research in Science Teaching, 39(3), 185–204.
Peirce, C. S. (1978). [Harvard] Lectures on pragmatism. Collected Papers of C. S. Peirce (pp. 171–172). Cambridge, MA: Belknap Press of Harvard University Press.
Pyle, E. (2008). A model of inquiry for teaching earth science. Electronic Journal of Science Education, 12(2), 3.
Raup, D. (1999). The nemesis affair: A story of the death of dinosaurs and the ways of science. New York: WW Norton.
Roth, W.-M., & Bowen, G. M. (2001). "Creative solutions" and "fibbing results": Enculturation in field ecology. Social Studies of Science, 31(4), 533–556.
Rudolph, J. L. (2003a). Some thoughts on portraying epistemology in today's classrooms: A reply to Garrison. Science Education, 87(1), 90–93.
Rudolph, J. L. (2003b). Portraying epistemology: School science in historical context. Science Education, 87(1), 64–79.
Rudolph, J. L. (2007). An inconvenient truth about science education. The Teachers College Record. Published February 9, 2007, on the Teachers College Record website, ID Number: 13216. Retrieved June 27, 2013, from http://www.tcrecord.org/.

Rudolph, J. L., & Stewart, J. H. (1998). Evolution and the nature of science: On the historical discord and its implications for education. Journal of Research in Science Teaching, 35(10), 1069–1089.

Schumm, S. (1998). To interpret the Earth: Ten ways to be wrong. Cambridge, England: Cambridge University Press.

Sibley, D. F. (2009). A cognitive framework for reasoning with scientific models. Journal of Geoscience Education, 57(4), 255–263.

Smoot, G., & Davidson, K. (2007). Wrinkles in time: Witness to the birth of the universe. New York: Avon Books.

Sober, E. (1993). Philosophy of biology. Boulder, CO: Westview Press.

Traweek, S. (1992). Beamtimes and lifetimes: The world of high energy physicists (pp. 187). Cambridge, MA: Harvard University Press.

Tucker, A. (2011). Historical science, over- and underdetermined: A study of Darwin's inference of origins. British Journal for the Philosophy of Science, 62(4), 805–829.

Turner, D. D. (2007). Making prehistory: Historical science and the scientific realism debate. Cambridge, England: Cambridge University Press.

Turner, D. D. (2013). Historical geology: Methodology and metaphysics. Geological Society of America Special Papers, 502, 11–18.

Walton, D. (2004). Abductive reasoning. Tuscaloosa: The University of Alabama Press.

Weiner, J. (1995). The beak of the finch: A story of evolution in our time. New York: Vintage Books.

Whewell, W. (1837). History of the inductive sciences: From the earliest to the present times. London: Cass.

Science
Education

Epistemology, Sociology, and Learning and Teaching in Physics

CRISTINA SIN
Centre for Research in Higher Education Policies (CIPES), 4450-227 Matosinhos, Portugal

Received 13 February 2013; accepted 18 December 2013
DOI 10.1002/sce.21100
Published online 14 January 2014 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: This paper explores the relationship between epistemology, sociology, and learning and teaching in physics based on an examination of literature from research in science studies, history and philosophy of science, and physics pedagogic research. It reveals a mismatch between the positivist epistemological foundation which seems to underpin the teaching of physics at the undergraduate level and the tentative nature of knowledge and the primarily social-constructivist process of knowledge creation which characterise the practices of professional physicists. Attention is drawn to the consequences of neglecting this mismatch, which is detrimental to students' understanding of the nature of the discipline, their conceptual development, and the acquisition of skills essential not only for a scientific career but also for students' development as individuals and citizens. The paper argues for the explicit contemplation of disciplinary epistemology in physics teaching and in pedagogic research to improve student learning and for the avoidance of the dangers of epistemological essentialism. © 2014 Wiley Periodicals, Inc. *Sci Ed* **98:**342–365, 2014

SITUATING PHYSICS LEARNING AND TEACHING IN CONTEXT

The role of disciplinary characteristics, epistemological or sociological, in shaping higher education pedagogy has received little attention from researchers (Hativa & Marincovich, 1995; Krause, 2012; Kreber, 2009; Neumann, 2001; Trowler, Saunders, & Bamber, 2012; Ylijoki, 2000). Learning and teaching in higher education physics has been discussed by McDermott and Redish (1999), Redish (2003), Redish and Steinberg (1999), Tobias (1992), and van Heuvelen (1991), to name but a few. The American Institute of Physics's

Correspondence to: Cristina Sin; e-mail: csin@cipes.up.pt

conferences also provided a wealth of resources on physics learning and teaching (Engelhardt, Churukian, & Rebello, 2012; Redish & Rigden, 1996). In addition, the results of the pan-European Tuning project, advocating pedagogic approaches grounded in the development of student competences (Tuning Project, 2008), and the resources of the Physical Sciences Centre in the UK[1] have proposed innovative pedagogies for improved student learning. However, the relationship between pedagogy and epistemological aspects, that is, how epistemology informs (or can inform) learning and teaching, has been addressed to a lesser extent. Matthews (1997), for instance, discussed the rise of constructivism as a new pedagogical paradigm toward the end of the twentieth century and presented various viewpoints about how this approach reflected the epistemology of science. Some more concrete examples of the integration of epistemology in learning and teaching are found in an account of how technoscience (a reconceptualized epistemological foundation for physics through its unification with technology) can be used to improve teaching and learning (Tala, 2009); a parallel between the epistemology of modeling (which illustrates the transition from abstract to concrete) and science teaching (Sensevy, Tiberghien, Santini, Laubé, & Griggs, 2008); a reconstruction of the epistemology of experiments with positive effects for students' learning and their own construction of knowledge (Koponen & Mäntylä, 2006); and a theoretical framework for an epistemological modeling of teaching-learning sequences that draws on studies of scientific practice (since understanding science implies some understanding of the practices involved in scientific inquiry; Psillos, 2004).

Considering pedagogy in relation to disciplinary epistemology seems to invite the harmonization of learning and teaching in physics with the nature of knowledge and the process of knowledge creation characteristic of the discipline with a view to enhancing student learning. Such approaches could be seen as responses to traditional instruction and the knowledge transmission model still prevalent in the teaching of physics (DeHaan, 2005; Redish & Steinberg, 1999; Thacker, 2003). A survey in the United States found that only a minority of students engaged with active learning or real-world problem solving in their introductory science courses; in the majority of cases, the typical practice was the lecturer delivering information (DeHaan, 2005). This conventional pedagogy came under further criticism that it was not effective in developing students' understanding, with some urging reforms in science education in general, physics included (DeHaan, 2005; National Science Foundation, 1996; Redish & Steinberg, 1999; Taylor, Gilmer, & Tobin, 2002; Tobias, 1992). Calls have been made to move away from lectures and, building on constructivist principles, to provide increased opportunities for students to discuss the nature and content of disciplinary knowledge. Having started to exert influence on science education at the end of the twentieth century, constructivism purported that meaning making takes place during students' interaction with the environment and advocated active experience with the physical world (Matthews, 1997).

Reform attempts in physics pedagogy have gained expression in what has come to be known as "physics education research," spurred on by gaps identified between instructors' expectations of student learning outcomes and actual conceptual understanding. However, physics education research has gone beyond highlighting the shortcomings of traditional instruction, giving rise to examples and proposals of innovative pedagogic methods (Heron & Meltzer, 2005). Indeed, physics has been a pioneering discipline in pedagogic improvement (DeHaan, 2005). In this respect, a comprehensive review of advances in classroom physics (Thacker, 2003) noted that curricula and courses had been redesigned with increased attention to conceptual understanding and the cognitive skills required to understand and apply physics concepts, attractive teaching environments and situations (such as "real-life"

[1] http://www.heacademy.ac.uk/physsci/, accessed August 2, 2013.

applications, hands-on environments, teaching modern physics and quantum mechanics concepts earlier in the curriculum), interactive engagement of students, and the use of technology. Concrete suggestions and examples of alternative pedagogical methods abound in the literature: problem-based learning as a "powerful alternative" to the passive lecture in introductory courses (Allen, Duch, & Groh, 1996); "interactive engagement" strategies, claimed to be more effective than traditional passive methods in enhancing students' understanding in conceptually difficult areas (Hake, 2002); enhancement of students' learning through participation in classroom demonstrations as opposed to acting as passive observers (Crouch et al., 2004); a grounded theory for students' construction of knowledge including talk and writing strategies to facilitate understanding of science concepts (Syh-Jong, 2007); and a design of teaching sequences based on a social constructivist perspective of learning consisting of three phases (staging the scientific story, supporting student internalisation, and handing-over responsibility to the students; Leach & Scott, 2002). These are only a few illustrative examples; documenting extensively the efforts to innovate physics pedagogy lies outside the scope of this paper. Yet, despite advances in the teaching of physics, there has been no wide-ranging progress in the way university courses are taught at most institutions. Instead, changes have been local, specific to a university or to a particular professor (Thacker, 2003). This is the reason why, when discussing pedagogy, the focus of this paper lies on traditional instruction methods, while acknowledging the recent developments in the context of physics education research.

Against this backdrop, the paper seeks to bring loose ends together and explore the relationship between disciplinary epistemology, sociology, and pedagogy, with a view to understanding how this relationship influences student learning at the level of curriculum content, knowledge transmission and acquisition, conceptual understanding, generic skills[2] development, assessment, research training, and so forth. It argues for the integration of epistemological and sociological considerations in the teaching of physics, further to observed disparities between the process of science making and the social constructivist practices of professional physicists, on the one hand, and undergraduate pedagogy generally informed by a positivist epistemology, on the other. Disregarding this mismatch, it is argued, has negative consequences for students' understanding of the nature of the discipline, their conceptual development, and the acquisition of skills that are essential not only for a scientific career but also for students' development as individuals and citizens.

The paper starts by delineating some key concepts—epistemological essentialism (Trowler, 2013), classification, and framing (Bernstein, 1971)—that serve to describe disciplines and disciplinary practice and can help articulate the link between disciplinary epistemology, sociology, and pedagogy in physics. Next, these three dimensions—epistemology, sociology, and pedagogy—are analyzed in turn. First, the paper discusses the existence of conflicting epistemologies (a positivist view of physics versus a social-constructivist, relativist one) both from a history and philosophy of science perspective and based on scientists' views of the nature of science explored by research in science studies. Second, in moving on to the sociology of physics, it is noted how its sociological aspects support a social constructivist epistemology. Sociology is explored with respect to three dimensions: scientific activities that result in knowledge creation and validation, social patterns of interaction among physicists, and wider societal issues related to the underrepresentation of some social groups in physics. The sociological insights acquire relevance because they

[2]Generic skills or attributes are understood here as student skills or attributes assumed to transcend the disciplinary context, transferable from one context to another, for example, critical thinking, problem solving, and communication (Jones, 2009b). However, Jones reconceptualized these as "discipline knowledge in action," an expression of the relationship between knowledge and the world, the application of knowledge to theoretical or practical problems, and the organized expression of that understanding.

challenge traditional views on the nature of science based on a positivist epistemology. Third, attention turns to learning and teaching and the relationship with the epistemology and sociology of physics discussed in the preceding sections. The analysis reveals a mismatch between contemporary views on the nature of science and the process of knowledge production of a mainly tentative nature and pedagogical choices mostly based on positivist principles, showing a variety of aspects in which this inconsistency can be detrimental to student learning and development. Finally, the paper concludes with a synthesis of the insights gained and makes some recommendations for pedagogic practice: first, the incorporation of epistemological and sociological considerations in learning and teaching to better reflect the evolution of the physics knowledge corpus and professional physicists' practices of knowledge creation and, second, the rejection of disciplinary essentialism based on positivist views of science in teaching, which fails to develop competent scientists and critical, discerning individuals.

PHYSICS SEEN THROUGH THE CONCEPTS OF DISCIPLINARY ESSENTIALISM, CLASSIFICATION, AND FRAMING

Epistemological Essentialism

This study draws on the literature that addresses the characteristics of physics knowledge (epistemology), the sociological and social aspects of physics (sociology), and learning and teaching (pedagogy). In bringing together these three dimensions and analyzing their interconnectedness, the paper distances itself from epistemological essentialism (Trowler, 2013), that is, a deterministic relation between knowledge characteristics of a discipline and academic practices. Epistemological essentialism stresses the homogeneity of specific disciplinary features, acting as unique identifiers that mark each discipline as being itself. It also bestows upon disciplines generative power, that is, their essential knowledge properties are claimed to generate, directly and universally, specific characteristics and practices among disciplinary practitioners, including at the level of pedagogy (Trowler, 2013).

Such an example of epistemological essentialism is provided by the description of physics as a hard/pure discipline in Becher and Trowler's (2001) work on academic tribes and territories. Physics is argued to be hard (vs. soft) on account of its clear paradigm, i.e., the consensus among the discipline's constituency on its epistemological territory and the methods of knowledge production, and pure (vs. applied) on account of its focus on theoretical knowledge rather than practical knowledge application. As a hard/pure area, it is described as follows:

> cumulative; atomistic (crystalline/tree-like); concerned with universals, quantities, simplification; impersonal, value-free; clear criteria for knowledge verification and obsolescence; consensus over significant questions to address, now and in the future; results in discovery/explanation. (Becher & Trowler, 2001, p. 36)

Nevertheless, epistemological essentialism has been challenged by theories that acknowledge that a multiplicity of factors, for example, social and individual ones, influence learning and teaching practices (Trowler et al., 2012). Social-constructionist theories argue that disciplines contain several narratives, constructed in specific contexts, shared and developed over time (Lindblom-Ylänne, Trigwell, Nevgi, & Ashwin, 2006; McCune & Hounsell, 2005), whereas individual agency theories suggest that individuals, through belief, decision, and action, shape disciplinary structures and practices (Hativa & Goodyear, 2002). Krause (2012), too, argues that traditional territories and tribal boundaries are becoming

increasingly blurred, noting variations in the sense of belonging to disciplinary teaching communities. Therefore, in looking at physics epistemological features and their influence on pedagogy, this paper does not suggest that these are exclusive determinant factors. Instead, it will challenge the essentialist description purported by Becher and Trowler (2001), drawing on evidence from the history, philosophy, and sociology of science and from research in science studies.

Classification and Framing

Classification refers to the separation and the strength of boundaries between the contents of discrete knowledge areas; it can be strong when areas are "well insulated from each other by strong boundaries," or weak when the insulation between content is reduced because the boundaries between them are blurred (Bernstein, 1971, p. 49). Whereas classification, in dealing with knowledge areas, has relevance for epistemology, framing describes pedagogy and the imparting of educational knowledge. *Framing* characterizes teachers' and learners' degree of control over the selection, organization, and pacing of the knowledge transmitted and received in the pedagogical relationship with respect to the options available to teachers and students. When strong, it entails reduced options; when weak, it entails a range of options (Bernstein, 1971, p. 50).

Wide agreement over what constitutes the core knowledge of physics (Becher, 1990; Becher & Trowler, 2001; Cole, 1992; Kekäle, 1999)—also conveyed by the concept "community consensus knowledge base" (Redish, 1999)—would suggest that physics is a strongly classified discipline, according to Bernstein's classification concept. Yet, consensus does not apply to science in the making, since at the research frontier competing theories dispute what nature's laws are (Cole, 1992). Regarding the strength of boundaries, Becher (1990) lists some limited overlap between physics and engineering (solid-state materials) and physics and biology (the structure of proteins), as well as between theoretical physics and mathematics. However, these are deemed to be exceptions, contrasts being clear overall. Nonetheless, with the sophistication of knowledge, disciplines have become increasingly intertwined and an "extraordinary confluence of disciplines" (Galison, 1996) has taken place since the mid-twentieth century. The simulated realities in the Monte Carlo experiments are a telling example: "part of mathematical statistics and yet often classified as part of physics ... not quite pure mathematics, not quite just part of nuclear weapons design, yet perhaps, simultaneously both these and more" (Galison, 1996, p. 15). Another example of confluence is the integration between technology and science in physics experimentation, rendered by the concept "technoscience" (Tala, 2009), to capture the unifying view of physics and technology in light of the cognitive role of technology in knowledge construction through experimentation. The increasingly interdisciplinary nature of scientific endeavors, therefore, implies perhaps a tendency toward a weaker classification of physics than that suggested by some scholars. Connections with other disciplines have multiplied and strengthened through more frequent interdisciplinary research and teams working together.

Moreover, Bernstein (1971) highlights the acute sense of identity and community belonging encountered in classified knowledge areas. However, physics does not appear homogeneous and conflict-free, containing divisions despite apparent unity. Various cultures and traditions exist within physics, meeting around "trading zones," in continuous transformation (e.g., the changes brought by the advent of the computer in the physicists' work and identity), but whose overlap has been essential to the discipline's continuity and evolution (Galison, 1997). In fact, disagreements appear to have favored advances in the field: Its evolution has not been "a smooth striding forth, but a survival of errors, a series of

revolts and revolutions, and thus also a history of forgetting and suppression" (Lepenies, 2006). The development of science not through a linear evolution whereby one theory builds upon another, but through fractures, with one paradigm replacing a previous one (Galison, 1997; Kuhn, 1962; Lepenies, 2006), represents another argument for the rather weak classification of physics at its knowledge frontiers.

However, core knowledge does enjoy consensus and this translates into curricular coherence at the undergraduate level (Cole, 1983; Kehm & Eckhardt, 2009). As supported by a survey of 152 physics bachelor programs (Kehm & Eckhardt, 2009), undergraduate curricula are rather similar in different European countries aiming to build a foundation of physics knowledge and methodologies and, thus, are illustrative of strong classification. The survey found that the first 2 years of a bachelor program in physics tended to be similar everywhere, "because students have to be familiarized with the tools of the trade and the subject matter." The third year of the program was usually dedicated to project work enabling a certain degree of specialization (Kehm & Eckhardt, 2009, p. 18). As to framing (Bernstein, 1971), likely because of existing consensus over core knowledge and what students should cover, this appears relatively strong in undergraduate education, manifest in the selection and organization of knowledge (Becher, 1990; Cole, 1983; Kehm & Eckhardt, 2009). However, weaker classification and framing were found in graduate programs, which were characterized by specialization and a pronounced research orientation (Kehm & Alesi, 2010). This gives more control to academics over the programs' direction and to students over their specialization.

The next sections discuss the epistemological and sociological aspects of physics, followed by an analysis of their expressions in teaching and learning. Equipped with the insights gained during the analysis, the paper then returns to the concepts of disciplinary essentialism, classification, and framing and their relevance to the appreciation of the relationship between the three dimensions (epistemology, sociology, and pedagogy).

CONFLICTING EPISTEMOLOGIES

Epistemology as a subfield of philosophy is concerned with knowledge, specifically *what* we know and *how* we know it. Hofer and Pintrich (1997) refer to these two dimensions as the *nature of knowledge* (what one believes knowledge is) and the *nature or process of knowing* (how one comes to know). These dimensions represent our reference in the examination of physics knowledge and the methods for its creation and validation.

Positivism Versus Constructivism

Becher and Trowler's (2001) description of physics as a hard/pure discipline, presented earlier, denotes a vast disciplinary area preoccupied with uncontroversial, context-free knowledge, whereas the process of knowing is characterized by objectivity, discovery, and logic. It appears to lean excessively on a positivist epistemology of physics. However, alternative claims from the history and philosophy of science and findings from research in science studies suggest that Becher and Trowler's essentialist depiction might need reviewing. For example, social constructivism, a perspective in the sociology of science which gained momentum in the past decades of the twentieth century, claims that it is not nature's laws that determine the intellectual content of science, but that science is socially constructed in the laboratory by scientists and that local contextual conditions shape scientific practice (Brannigan, 1981; Cole, 1992; Fine, 1996; Latour & Woolgar, 1986; Pickering, 1984). A powerful metaphor to suggest the man-made, subjective nature of science is the golem, a Jewish mythology creature "of our art and craft," "a humanoid made by man from

clay and water" (Collins & Pinch, 1993, pp. 1–2). Similarly, science studies have challenged traditional claims that science is value-free and universal and have contextualized science historically and culturally. Our representations of the world at any point in time are but "stations along the chain of experience," which through successive rectifications lead to revised versions (Latour, 2008). For Latour, time, rectification, instruments, people, and institutions are the "very stuff" of science. Thus, in looking at the dynamics of scientific work and how knowledge claims emerge from scientific practice—molded and constrained by cultural norms and values, organizational and institutional structures, economic and political power relationships, interests, and so on—science studies have emphasized the sociocultural dimension of scientific knowledge construction (Collins & Pinch, 1993; Galison & Stump, 1996; Knorr-Cetina, 1995; Stump, 1996), an aspect to be dealt with further in the section Sociocultural Aspects of Physics.

Such conflicting views about the nature of knowledge and its creation process suggest the existence of physics epistemologies at odds with each other. According to a positivist view, the knowledge corpus of physics consists of objective natural laws. But according to constructivist views, these are socially constructed artifacts. In the following, practicing scientists' views of the nature of science are briefly explored to get a perspective from disciplinary "insiders."

Practicing Scientists' Epistemologies

The existence of parallel epistemologies can be explained through the historical evolution of the views of the nature of science. In physics, epistemologies have changed over time through the shift from a classical, deterministic approach to a quantum, indeterministic conceptualization of the discipline (Abd-El-Khalick & Lederman, 2000). However, although epistemological views appear situated primarily in a historical context (as discussed in the section Epistemologies Among Educators and Implications for Pedagogy), both positivist and constructivist positions are still encountered among practicing scientists. On the one hand, vehement arguments deny that scientific truth should be relative to a given local and social framework (Kragh, 1998). According to such positivist opinions, unexpected discoveries (e.g., Rontgen's discovery of rays) or quantitatively precise and confirmed predictions (e.g., the discovery of Neptune) act as evidence that objects or phenomena exist in the natural world. Therefore, although discovery is a social process, discovered objects are "parts of nature and cannot be negotiated away if the scientists should so decide" (Kragh, 1998, p. 6).

Positivist views were also revealed by a study into the views of the nature of science of 24 scientists from various disciplines (Schwartz & Lederman, 2008): Nine of these suggested either that science attains certain absolute knowledge or that science progresses nearer and nearer to certain knowledge through pure discovery, dismissing interpretation as unnecessary. While finding variation among scientists' views, no overarching pattern was noted to suggest a predictable relationship between discipline and expressed views. At the opposite end, some prominent scientists' accounts on their views of knowledge and science (Wong & Hodson, 2009, 2010) indicate a belief that scientific theories are human constructions, created, sustained, and modified through social processes. However, for these scientists, scientific knowledge goes beyond being a mere social construct; at the same time, they believe in the rationality of science, and all view it as true everywhere, at least in relation to established knowledge (what we earlier referred to as the disciplinary consensus over core knowledge). In a similar vein, another study (Yore, Hand, & Florence, 2004) found that some scientists held "evaluativist" views and rejected absolutist or relativist extremes. They described science in terms of arguments, hypothesis testing, or tentative science. Among

physicists, a consensual epistemological adherence does not appear to be shared either, as testified by Barad (2007), Galison and Stump (1996), and Pickering (1995). In an acute form, this is demonstrated by the disagreements about the epistemological interpretation of quantum mechanics (Cross, 1991; Freire, 2003).

A fact to bear in mind, however, is that practicing scientists usually do not ponder consciously their epistemological stance, but concentrate instead on their everyday practice. "Privileged access" to what their practice entails does not imply a similar level of access to its epistemological underpinnings (Abd-El-Khalick, 2011). This invites the consideration of other sources of evidence, such as the scientific process of knowledge creation, to get further insight into the discipline's epistemological foundations. The process of science making in the laboratory has been the object of microsociological studies of science, which will be discussed next alongside other sociocultural aspects of physics.

SOCIOCULTURAL ASPECTS OF PHYSICS

In investigating how science is practiced and constructed in society, the sociology of science lays emphasis on its human and societal component, questioning its apparently "mythical" status (Cunningham & Helms, 1998). The following discussion addresses sociological aspects of science (and implicitly physics) and their subsequent implications for epistemology. In science studies, these aspects are tackled at microsociological and macrosociological levels (Cunningham & Helms, 1998). The next sections dwell on these, as well as on the social patterns of interaction within the physics community. The combined implications of epistemology and sociology for pedagogy are explored in the section Pedagogy: Epistemological and Sociological Expressions.

Microsociological Studies: Scientific Practice

Microsociological studies zoom in on the everyday practices of scientific production, offering depictions of the knowledge creation enterprise as it takes place in laboratory settings. They analyze how scientific undertakings and scientists' interactions and ways of working lead to the generation of scientific claims; how evidence is evaluated and negotiated in the scientific community; and how scientific knowledge gains validation and acceptance (Collins & Pinch, 1993; Gooding, 1990; Knorr-Cetina, 1995). Minute attention to the processes of knowledge creation has raised epistemological questions in relation to the unbiased nature of science and the supremacy of the scientific method in the production of irrefutable knowledge. Contradicting the objectivity of science, such studies have revealed the imprint of individual and cultural aspects and values on the process of knowledge production, justification, and its outcomes. Social aspects have thus become difficult to "bypass," and epistemology has become intertwined with sociology (Tala, 2009).

For instance, studies have documented the disparity between the messy research process and the linear accounts of science presented in published material (Gooding, 1990; Wong & Hodson, 2009). The latter leave out or play down the "messiness" of empirical work, concealing the extent to which scientists' accounts are "reconstructions rather than records." Reconstruction emerges as part and parcel of the scientific endeavor, whereby scientists "iron the reticularities and convolutions out of thought (and action) to make a flat sheet on which a methodologically acceptable pattern can be printed" (Gooding, 1990, p. 5). Similarly, according to a physicist's opinion in a study by Wong and Hodson (2009), the process and method of scientific investigation is flexible, chaotic, and needing creativity and imagination in all the stages of inquiry. The positivist appearance of scientific results thus contrasts with the less-positivist nature of scientific practice.

As to the evaluation of claims, analyses of scientific practices suggest that the consensus of the scientific community acts as enforcer of the validity of evidence and methodology (Cole, 1992; Tala, 2009; Wong & Hodson, 2010). Without a community structure, the justification process would result in "endless regression" and no "conclusive views" (Tala, 2009). However, there are different views on the extent of social manipulation: some claim that knowledge becomes authoritative through social institutional power, with the winner of the controversies invoking the idea of nature and imposing the rules of future research, whereas others merely acknowledge "the rather indisputable fact that the scientific inquiry is a social process and the reasoned judgment is itself socially defined" (Tala, 2009, p. 279). Thus, according to positivism the rigor of the scientific method separates justified belief from mere opinion, whereas social studies of science point out the community consensus as the arbiter of justified belief.

Social Interaction Patterns and the Pride-of-Place of Research

As a dimension related to the process of knowledge creation in physics, physicists' social interaction patterns deserve attention too, especially because of their reflection (or lack of) in pedagogic practice, as discussed later. It is argued here that the prime driver and molder of social interactions in physics is research as the practice that generates knowledge. Therefore, as a central component in physicists' activities, research assumes a "pride-of-place" position.

Several studies identify the strong research orientation and the tight research organization as defining features of physics (Becher, 1990; Becher, Henkel, & Kogan, 1994; Hermanowicz, 2006; Smeby, 1996, 1998, 2000). Research represents a critical element of physicists' career, capable of making the difference between success and failure, and steers their social behavior. Therefore, the qualities that physicists consider essential for career success invariably revolve around research (Hermanowicz, 2006). Persistence emerges as a paramount quality, as physicists deal with rejection throughout their working life. Peer reviews of papers and grant proposals often fail to yield results, as does the process of experimental and theoretical work (Hermanowicz, 2006). Smartness and civility, understood as collegiality that contributes to a work environment conducive to productive research, are other essential qualities for physicists. So is ruthlessness, related to the research endeavor and persistence, to "picking time to work on things" and to publishing, since well-known physicists are famous "because when a new idea comes out, they are quick about writing a paper on it, even if it's half-baked" (Hermanowicz, 2006, p. 143).

The tight research organization and the "ruthlessness" linked to research ambitions seem to result from the people-to-problem ratio and the urban character of physics (Becher & Trowler, 2001, pp. 106–108). Making an analogy with ways of life, Becher and Trowler classify disciplines into urban and rural: narrow areas of study clustered around a few prominent topics, versus broad stretches of intellectual territory with vaguely delimitated problems and a variety of themes. In contrast to rural areas that display rather individual endeavors in settings with little interest overlap, in physics teamwork, collaboration and competition are common social practices, essential to speed up knowledge generation, extend expertise, and validate and reject claims (Ford, 2008; Wong & Hodson, 2010). The intense competition generates a concern with rapid publication (Becher, 1990; Becher & Trowler, 2001; Hermanowicz, 2006; Wong & Hodson, 2010). Associated with the indispensable interaction with colleagues and the desire to keep up-to-date are networking, the common circulation of articles before publication, and frequent participation in conferences (Becher, 1990). The pivotal role of research becomes evident again. It is through the medium of research that the apparent contradiction between ruthlessness and physicists' sociability

could be explained. Both are necessary for the advancement of knowledge. Ruthlessness, applied to oneself and one's own time, enables progress in research and dissemination, but at the same time socializing and networking are indispensable to test ideas and get new insights.

The sports metaphors proposed by Kekäle (1999) are suggestive of the social relationships in urban and rural disciplines. In physics, the sense of collective concerns and collaboration prevails: It is like a fast team sport, researchers working together and competing intensely against other teams. In contrast, rural fields such as history are like jogging: people participate on their own or in small groups, the distance between start and finish is relatively long, the speed is slow, and there are many interesting paths to follow, so participants might not stay on the same track and reach the same destination (Kekäle, 1999, pp. 233–234).

Macrosociological Studies: Science and Societal Issues

Engagement with physics and its disciplinary community is not experienced equally by all those involved (both existing and potential members), as testified by feminist critiques of science and postcolonial science studies. As examples of macrosociological studies, these tackle the relationship between science and society by investigating how issues such as power, politics, race, religion, or gender interact with science. More specifically, such studies have revealed the existence of barriers for certain social groups, looked into the causes of discrimination and questioned conventional understandings of the nature of science.

One line of research has analyzed the participation of women and ethnic minorities in science, highlighting the discrimination and stereotypes that these groups encounter in gaining equal access to science, in proving that they can do science, in gaining resources once they have become members of the scientific community, or in getting equal recognition for their achievements (Blickenstaff, 2005; Carlone & Johnson, 2007; Etzkowitz, Fuchs, Gupta, Kemelgor, & Ranga, 2008; Harding, 1991; Nelson & Brammer, 2010; Rosser, 2012; Tyson, Lee, Borman, & Hanson, 2006). Other studies have investigated the reasons for discrimination, i.e., the ethnocentric and androcentric nature of science which has led to the marginalization of women and ethnic minorities. Scholars have revealed the gendered and white image of science (Harding, 2008; 2009), including in physics. An examination of the literature on gender and physics pinpoints the generally unwelcoming workplace culture for women, inverting the source of concern from the "problem of women in physics" to the "problem of physics with women" (Götschel, 2011). Postcolonial science studies, in turn, have questioned the supremacy of White Western science, claiming the equal status and worth of indigenous knowledge systems (see, e.g., Carter, 2008; Harding, 1998; Paty, 1999; Seth, 2009). On account of science being perceived as synonymous with the epistemologies and practices of the developed world, Western Science has been referred to as the "ethnoscience" (Harding, 1998), which has subjugated other non-Western scientific and cultural traditions. Therefore, an inclusive and multicultural view of science is advocated that acknowledges local systems of knowledge—previously dismissed as unscientific—as attempts to make sense of the natural world in response to local needs (Carter, 2008; Harding, 2009).

Therefore, and of particular relevance here, macrosociological studies—both postcolonial, as above, and feminist (Mayberry, Subramaniam, & Weasel, 2001; Subramaniam, 2009)—have also challenged conventional understandings of the nature of science. They have raised epistemological questions about the nature of scientific knowledge, the way in which science is conducted, and the fundamental assumptions upon which it is based. Feminist science studies, for instance, have engaged in a "cultural deconstruction of

science" (Bartsch, 2001) and recognized the interdependency of "natures and cultures" (Mayberry et al., 2001). Questioning science's claims of neutrality, such writings suggest that there are no objectively knowable facts, arguing for an understanding of science as a socially and culturally determined set of practices. Feminist theories in physics have also changed its image from "an area of eternal truth and solid knowledge" to one of "human endeavour and processes of solidification" (Götschel, 2011), as illustrated by Barad's (2007) theory of agential realism, which acknowledges the entanglement of natures and cultures.

This section has dwelt briefly on sociological perspectives of science and physics, both at the microlevel as regards the production of scientific knowledge in the laboratory and at the macrolevel in the relationship between science and societal issues of gender and race. Of significance to this paper, such sociological insights have exposed epistemological foundations of the nature of science and physics that challenge positivism.

PEDAGOGY: EPISTEMOLOGICAL AND SOCIOLOGICAL EXPRESSIONS

Epistemologies Among Educators and Implications for Pedagogy

A wealth of research has investigated the views about science held by educators (mostly at the preuniversity level) and students, in a variety of geographical contexts (Abd-El-Khalick, 2011; Abd-El-Khalick & Lederman, 2000; Belo, 2013; Iqbal, Azim, & Rana, 2009; Lederman, 1992; Lee & Witz, 2009; Tsai, 2006, 2007). The findings of these studies suggest that science educators often adhere to a positivist epistemology, in contrast with the views on the nature of science promoted by science education organizations that have undergone a constant evolution (Abd-El-Khalick & Lederman, 2000). As explained by Abd-El-Khalick & Lederman (2000), during the early 1900s the nature of science was associated with "The Scientific Method." Then, whereas the 1960s still emphasized inquiry and procedural skills, in the 1970s scientific knowledge started to be viewed as tentative, subject to change, probabilistic rather than absolute, resulting from human endeavors to make sense of nature, particular to historical contexts, and empirical. In the 1980s, the role of human creativity in elaborating theories and the social dimension of science started to be acknowledged. The 1990s continued to emphasize the historical, tentative, empirical, and well-substantiated nature of scientific claims, as well as the interaction between personal, societal, and cultural beliefs in the generation of scientific knowledge. However, in spite of these developments, a significant proportion of teachers still believed that scientific knowledge was not tentative or held a positivist, idealistic view of science (Lederman, 1992).

It is most likely a consequence of such views that knowledge transmission and students' systematic accumulation of factual information still appear to underpin to a large extent curricula and pedagogy in science in general and physics in particular (Duschl & Osborne, 2002; Lattuca & Stark, 1994; Neumann, Perry, & Becher, 2002; Smart & Ethington, 1995; Thacker, 2003; Wieman, 2007). Lattuca and Stark (1994) noted that, in hard fields, pedagogy at the undergraduate level is characterized by "curricular coherence," which means that students learn by building blocks of the discipline one upon another until reaching the prescribed level of understanding. In contrast, softer fields display curricular diversity, and knowledge is usually acquired by recursive patterns of research rather than by systematic accretion, using multiple perspectives and pursuing knowledge in several directions simultaneously (Lattuca & Stark, 1994, p. 419). Similarly, an analysis of course content in various disciplines (Donald, 1983) revealed differences. In social sciences, learning

occurred around clusters of loosely structured concepts where certain key ones acted as "pivots" or "organizers." In contrast, physics displayed hierarchical learning patterns with interlinked, tightly structured key concepts and with branches from more to less important concepts, suggesting an "all-or-none learning pattern" (pp. 37–38). The consequences of a highly prescriptive and tight curriculum, revealing a strong classification and framing (Bernstein, 1971) of undergraduate physics, will be addressed in the remainder of this section.

Knowledge Acquisition of "Ready-Made Science"

A perceived necessity of the "all-or-none learning" of "ready-made" scientific facts is probably what lies at the root of the emphasis on subject matter knowledge and on familiarity with "the foundations of the scientific canon" (Duschl & Osborne, 2002). Nevertheless, pedagogic practices based on the assumption of a vast, orderly knowledge area that students are supposed to assimilate systematically contradict the process of knowledge creation in physics, which was shown to involve messiness and collective and individual reconstruction. Positivist teaching approaches, in their varied manifestations, conceal the epistemic properties of scientific practice revealed by microsociological studies of science. Curricular material tends to hide the people and social contexts involved in the construction of science. Even when students are engaged in active scientific inquiry, there is often a push toward one right answer that promotes a singular vision of science (Barton & Yang, 2000). Thus, the image of the scientific process presented in science textbooks dismisses creativity as unnecessary, implying that dispassionate and systematic analysis of data will lead to secure conclusions (Wong & Hodson, 2009). In addition, classrooms are hierarchically structured, with the teacher and the text controlling which knowledge counts (Barton & Young, 2000; Cunningham & Helms, 1998; Duschl and Osborne, 2002), again indicative of the presence of strong framing and weak choices for students (Bernstein, 1971). Combined, such practices promote "scientific concepts over scientific contexts" (Barton & Yang, 2000), engendering a vision of science as factual, decontextualized, linear, objective, rational, and uncontentious, where learning becomes equivalent to retention of information (Barton & Yang, 2000; Neumann et al., 2002). The emphasis lies on "ready-made science" (with implicit messages about certain knowledge obtained through the scientific method), as opposed to "science-in-the-making," which emphasizes social construction (Wong & Hodson, 2010).

Moreover, students are confronted with an apparently neutral process of validation of empirical evidence in the form of "the scientific method." There is hardly any place for the "awkward student" (Mody & Kaiser, 2008) who reaches the "correct" answer via non-common-sense methods, considers alternative interpretations and new ways of doing things, and thus constructs knowledge while learning. For example, in traditional introductory university physics courses, laboratory activities usually consist of verifying principles that have been learned in lecture, and completion of the laboratory simply requires following a set of rules to get to the end result. Students neither engage in discovery nor practice laboratory skills necessary in research or in higher level courses (Thacker, 2003). Therefore, such practices hardly reflect the reality of physicists' day-to-day undertakings and the processes whereby claims are made, justified, and validated through the consensus of the scientific community. Consequently, pedagogical methods centered on acquisition of certain, absolute knowledge, which neglect the process of knowledge production, fail to make students aware of key sociological aspects of the discipline and the ensuing epistemological implications related to how knowledge claims have come into being and achieved validation.

Student Epistemology and Conceptual Understanding

In addition, such pedagogical practices give students a false impression of the discipline, that it is made up of facts about an objective reality and grows through the neat, systematic accumulation of knowledge. This perception is not without consequences, since epistemological beliefs have been shown to influence student achievement (Hammer, 1994; Lising & Elby, 2005; May & Etkina, 2002; Ryder & Leach, 1999; Songer & Linn, 1991; Stathopoulou & Vosniadou, 2007). For example, Stathopoulou and Vosniadou (2007) found that if students see physics knowledge as simple and/or certain they will focus on "piecemeal" factual information to the detriment of conceptual understanding, since they will be likely to filter out tentative and controversial information that contradicts existing knowledge. In contrast, perceptions of physics knowledge as complex, uncertain, and evolving determine students to focus more on relationships and their change in time. Unsurprisingly then, pedagogic methods concerned with mere accumulation of factual knowledge have often been highlighted as counterproductive to deep learning and conceptual understanding of physics (Bernhard, 2000; Duschl & Osborne, 2002; Ehrlich, 2002; Linder, 1992; Redish, 1999; Thacker, 2003; Wieman, 2007). One suggestion to counterbalance this negative effect entails the reduction of the cognitive load, while at the same time helping students see the interconnections of taught concepts, which is expected to direct their reasoning away from "novice" to "expert" thinking (Wieman, 2007). As "novices," they see physics as consisting of isolated facts, unrelated to the world around them, which they learn by memorization; on the contrary, as "experts," they see physics as a coherent structure of concepts that describe nature. When emphasizing the learning of subject matter, instructors wrongly assume that expert-like ways of thinking will follow and students are therefore not helped to develop metacognition (Wieman, 2007). Therefore, one could presume the existence of a belief among instructors that, in order for students to develop conceptual understanding, all the knowledge imparted is necessary (manifest in the all-or-none learning pattern). However, such a cognitive overload can have the opposite effect. In addition, two generic skills appear to be affected by the poor development of conceptual understanding: problem solving and critical thinking.

Problem Solving and Critical Thinking

Problem solving entails "hypothesis development and testing, use of mathematical modeling to describe and analyze the physical world, and awareness of issues of precision and rigor" (Jones, 2009a, p. 181). Its centrality in physics is uncontested, and the development of problem-solving skills is integrated in teaching and practiced in classroom and laboratory work (Jones, 2009a, 2009b; Redish, 1999; Thacker, 2003; Wieman, 2007). However, whether the way it is taught encourages conceptual understanding and deep learning is questionable. The importance of conceptual understanding—rendered by the concept of "knowledge structures"—to problem solving in physics emerges from research which found that students with fragile knowledge structures and weak links between distinct parts may not be able to activate the knowledge necessary to solve a problem (Sabella & Redish, 2007). Yet, despite being the most important skill taught in the undergraduate years, problem solving appears dominated by superficial mathematical calculations and fails to engage students in deeper analysis (Redish, 1999), suggesting instead the mere memorization of formulas. Similarly, some of the "top" students with high scores on the quantitative problems were found to have very low scores on the conceptual aspects of the subject (Bernhard, 2000). Such findings suggest that although students do apparently develop problem-solving skills, this does not necessarily go hand in hand with

the development of strong "knowledge structures" and cognitive processes characteristic of expert physicists.

In addition, a tension is noted between the emphasis on content knowledge and generic skills such as critical thinking or communication, the latter overshadowed by the primacy of the former (Jones, 2009a). To cover what is perceived to be a vast knowledge domain, the early years in the study of physics are dedicated to the teaching of physical concepts and principles deemed fundamental (in the form of factual information), whereas in social science or humanities, personal opinion and critical thinking are integrated early on as fundamental skills to be cultivated (Jones, 2009a; Lattuca & Stark, 1994). A survey of European physics bachelor programs (Kehm & Eckhardt, 2009) revealed, nonetheless, that a large proportion (78%) integrated the acquisition of generic skills. The most commonly mentioned were English-language skills (in non-English-speaking countries), communication skills, and project management skills, sometimes "outsourced" to other departments or teaching and learning support centers (Kehm & Eckhardt, 2009, p. 16). While this might suggest an increasing concern with equipping students with abilities relevant to their rounded development and a future scientific career, it is worth noting that the report does not mention critical thinking. Given its essential presence in a physicist's skills set, as discussed next, the question "why" springs to mind.

Research points out that physicists generally recognize that evidence can only support theories and not provide definitive answers and absolute truths (Jones, 2009b; Schwartz & Lederman, 2008; Wong & Hodson, 2009, 2010). Scientists working at the research frontier do not know what the laws of nature are and can reach different solutions in trying to interpret these. Insufficient data lead to the coexistence of multiple theories and divergent views, with differences in interpretations eventually resolved by new evidence. Consequently, knowledge representations of the world undergo evolution (Latour, 2008) and much of what was accepted as true in the past is now believed to be wrong (Cole, 1992). In addition, creativity and imagination enter in the formulation of interpretations and theories (Wong & Hodson, 2009). These facts imply that uncertainty does belong in physics and that critical thinking represents an indispensable skill for physicists faced with the relativism of knowledge. Yet, uncertainty is concealed by teaching approaches when these are based on positivist epistemologies. Students cannot easily embrace a critical attitude in a field with apparently uncontested knowledge and clear criteria for knowledge verification. This might explain why critical thinking is perceived as a challenge to teach in undergraduate physics (Jones, 2009a). Even more worrying, students were found to hold more novice-like beliefs after having attended an introductory physics course than before it (Wieman, 2007). One can only guess that it was the nature of the curriculum and pedagogical approaches, suggesting the certainty of knowledge and the objectivity of its methods, which was responsible for a shift toward novice thinking rather than expert thinking. Thus, in a feminist perspective on the physics curriculum, Barad (1995) laments the "acritical–anticritical" pedagogy embraced by the physics community and argues for the teaching of the "uncertainty principle." Similarly, Feynman criticizes teaching approaches that generally follow one path and induce students to believe in the validity and uniqueness of the "fashionable" theory, rather than imparting to students a wide range of physical viewpoints:

> If every individual student follows the same current fashion in expressing and thinking about electrodynamics or field theory, then the variety of hypotheses being generated to understand strong interactions, say, is limited. Perhaps rightly so, for possibly the chance is high that the truth lies in the fashionable direction. But, on the off-chance that it is in another direction—a direction obvious from an unfashionable view of field theory—who will find it? (Feynman, 1965)

The realization that there are multiple theories usually occurs at the postgraduate level, once students start undertaking research and create new knowledge. Confronted with ambiguity, they develop critical thinking. Once again, a contradiction seems to emerge between physicists' practices, which involve constant searches and attempts to resolve uncertainties to make sense of nature, and the positivist teaching approaches, which present knowledge as uncontested facts and fundamental truths, hardly promoting an inquisitive, critical attitude toward the imparted knowledge, essential in a physicist's skills repertoire.

Assessment

The emphasis on objective content knowledge also appears to influence student assessment, hence the distinction between assessment based on memorization and the application of course material in hard sciences, as opposed to assessment that requires analysis and synthesis of course content and critical thinking in soft sciences (Braxton, 1995). Physics students pass courses by remembering facts and problem-solving recipes (Ehrlich, 2002; Wieman, 2007), which favors an impression that physics is effectively about memorization and the use of formulas. Again, such assessment practices ignore uncertainty as a dimension of physics epistemology and fail to develop students' critical inquiry abilities. Another noted tendency is that whereas the hard sciences give more weight to final examinations, soft fields show a tendency toward continuous assessment (Neumann, 2001). However, the above-mentioned survey of physics bachelor degrees (Kehm & Eckhardt, 2009) observed a change in the majority of continental European countries: a shift toward continuous assessment and a reduced emphasis on final summative examinations described as the typical, traditional examination method in continental Europe. Although the latter continue to have considerable weight, a recent concern with student-centered learning appears to have triggered a new practice, the assessment of learning outcomes after each module or unit of teaching. A large majority of survey respondents (60%) also reported that, in addition to knowledge, their bachelor programs assessed generic skills. Yet, these skills do not appear to include critical thinking. Therefore, assessment appears dominated by mastery of subject matter and mathematical formulas and fails to test students' development of capabilities indispensable to expert physicists.

Decontextualized Science: Effects on Underrepresented Groups

Besides the shortcomings identified so far, the image of science as objective, context-free, and unitary conveyed by curricula and pedagogic practice has the additional negative effects of alienation of women and minority groups. As it promotes a Western and gendered view of science (Harding 2008, 2009), underrepresented students find difficulty in relating to it, integrating it with their own contexts, and finding meaning in their learning. Their identities clash with the culture of science, leading to low participation, problematic integration, and frequent dropping out from science courses (Barton & Yang, 2000; Carlone, 2004; Carter & Smith, 2003; Jones, Howe, & Rua, 2000; Kozoll & Osborne, 2004; McCullough, 2004; Miller, Slawinski Blessing, & Schwartz, 2006). In physics, a literature review on gender and education (Danielsson, 2009) revealed pedagogical implications such as the duality of the student body in terms of student identities, with male students interested in the discipline for its own sake and female students struggling to relate physics, as it is taught, to their own reality.

Inspired by insights into the sociology of science, the science education literature offers several suggestions about ways to make pedagogy more inclusive and relevant to women and underrepresented groups. A self-evident method refers to inclusive curricular material

and textbooks that reflect gender and race diversity through accounts of the contribution of scientists from underrepresented groups and of indigenous sciences to scientific knowledge (Barton & Yang, 2000; Brickhouse, 2001; Snively & Corsiglia, 2001; Whiteley, 1996). Other methods, however, could potentially benefit the student body as a whole, beyond assisting the integration of underrepresented groups. They generally target the strong framing of educational knowledge (Bernstein, 1971) in the direction of handing over to students more options and control over their learning. For example, under the influence of feminist epistemology, feminist pedagogies challenge power relationships in teaching between instructor, subject matter, and students and promote instead a consideration of students' ideas and needs (Brickhouse, 2001). Such practices are likely to make science more attractive and engaging in general, while at the same time developing in students' high levels of scientific literacy. Concrete suggestions in this respect contemplate consideration of students' prior experiences of science and their interests (Barton & Yang, 2000; McCullough, 2004), interactive environments that promote cooperation and discussion in the classroom (Lorenzo, Crouch, & Mazur, 2006), and teaching not only the ready-made products of science but also knowledge about the processes of scientific production and the nature of science through engagement in activities that resemble scientists' practices (Cunningham & Helms, 1998; McGinn & Roth, 1999; Osborne, 2007). In addition, such approaches could have an added benefit: They could raise students' awareness of the subjective dimensions of science, the collective processes of knowledge creation and evaluation of evidence, the coexistence of conflicting theories, and the provisional character of knowledge, thus generating a more faithful alignment of pedagogy with the nature of knowledge and the process of knowledge production in physics. This alignment in fact occurs at the postgraduate level.

Research Training: Pedagogy in Tune

During research training at the postgraduate level, instruction finally seems to reflect the knowledge production and the social patterns of interaction characteristic of the physics community. Students' initiation to research, part of their formal training in postgraduate studies, does not seem to display the inconsistencies observed in undergraduate education. Instead, it appears to converge with the activities of expert physicists. The most likely explanation lies in the pride of place of research in the physics profession discussed earlier. In a university environment, this translates into the fact that physics academics identify themselves strongly with research, and less with teaching, and, as opposed to the arts and social sciences, they perceive supervision as research rather than teaching (Becher & Trowler, 2001; Moses, 1990; Smeby, 1996). They also spend large amounts of time on supervision, since students' work contributes to the department's research efforts. Smeby (2000) found that at the University of Oslo supervision time fluctuated significantly: 42 hours per year in the humanities and social sciences compared to 82 hours per year in the sciences.

Postgraduate students' integration in communities of practice reflects the tight organization of research and the urban nature of the discipline. Students work in a team alongside other students and staff who pursue similar research. They are often assigned topics directly associated with the supervisor's specialty (Becher, 1990; Smeby, 1998), and their work becomes part of the joint effort. In fact, physics academics believe that it would be difficult to do research in universities without graduate students—hence the mutual dependency in the relationship between staff and students, beneficial for both parties. Students get involved in real research, and staff have a genuine interest in the topic and progress since results will contribute to their own research (Smeby, 1998). A physics academic describes students as a resource and their contribution as positive: "they take part, solve problems and do a lot

of hard work" (Smeby, 2000, p. 59). Students' socialization into a community of practice is also evident in Ph.D. students' perceptions of research in different disciplinary areas. Whereas in medicine research is "a job to do," in the natural and behavioral sciences, students perceive research as a "personal journey." In the natural sciences, this journey includes learning how to be part of a research community (Stubb, Pyhältö, & Lonka, 2012). Thus, since a career in physics, within and beyond academia, is perceived to be intricately related to research, there is a pervasive concern among physics academics to train students in research skills (Sin, 2012). In soft and/or applied disciplines, one could also claim research to be a defining characteristic for the academic profession; however, it is less likely to be required for graduates who leave academia for industry.

Therefore, one can conclude that through involvement with research, postgraduate students get acquainted both with the uncertainty inherent in physics knowledge and with the complex process of knowledge creation and its social dimensions. It is only at the postgraduate level—already a springboard to the physics profession—that the tentative, socially constructed nature of scientific knowledge becomes obvious, testifying to a more faithful alignment between disciplinary epistemology, sociology, and pedagogy.

DISCUSSION AND IMPLICATIONS FOR PEDAGOGIC PRACTICE

The paper has set out to analyze the relationship between epistemology, sociology, and pedagogy in physics and has offered some examples of learning and teaching approaches and practices that illustrate a (mis)alignment with disciplinary epistemology and sociology. In so doing, it has raised questions about the disciplinary essentialism embodied in positivist epistemologies, warning that the assumption of the presence of some quintessential properties of physics (objective, logic, context-free, uncontroversial, etc.) can condition pedagogic practice in a way that is detrimental to students' understanding of the discipline, their learning, and their development. With such an insight, the paper casts doubt on the continuing authority of Becher and Trowler's (2001) characterization of hard/pure areas and, by extension, their clear-cut disciplinary classification that has informed much subsequent pedagogic research.

Coming back to the theoretical concepts of classification and framing (Bernstein, 1971), a dividing line becomes evident between undergraduate and graduate pedagogy. *Undergraduate teaching* appears to rely on a *strong classification*—clear knowledge boundaries that contain the *core physics knowledge*—and, deriving from it, to display a *strong framing* whereby instructors' and students' options with regard to selection, organization, and transmission of knowledge is limited. Strong classification and framing translate into a *tightly bound curriculum* that displays a resemblance across countries (Kehm & Eckhardt, 2009), suggesting the *universal and context-free* character of physics knowledge. The emphasis on the acquisition of this knowledge betrays a concern with *ready-made science*, that is, with *the outputs of the scientific process* of knowledge creation. In contrast, *postgraduate education* appears to be characterized by weak classification and weak framing. *Weak classification* reflects the *lack of consensus over frontier knowledge* and students, through research, get introduced to the *uncertainty* inherent in treading this knowledge territory. *Weak framing* is manifest in the range of choices available to students, since they have reached a level that entails *specialization* and decisions about research avenues worth pursuing. The preoccupation now lies in students' induction to authentic scientific practices of knowledge creation and validation through their integration in a research community, as well as in their socialization into the interaction patterns characteristic of the discipline. The emphasis is no longer on the output of the scientific process, but on *the scientific process itself*, or on *science in the making*. Undergraduate education thus appears to embrace a

positivist epistemology, whereas postgraduate education a relativist, social-constructivist one. It is the latter that is supported by evidence from research in science studies and by the history, philosophy, and sociology of science that suggest that the nature of scientific knowledge and the process of knowing are tentative, situated in a social and historical context and a result of individual and collective endeavors.

One can therefore argue that the strong classification and strong framing that characterize physics curricular knowledge and teaching at the undergraduate level are a consequence of an underlying positivist epistemology. Despite physics knowledge being documented to advance through radical shifts and disciplinary revolutions, its teaching appears to be characterized by tight organization, systematic assimilation of knowledge, and the "all-or-none learning pattern" (Donald, 1983). In addition, the emphasis on content knowledge hides from students the process of knowledge creation and its human and social dimension. Therefore, these pedagogic approaches give an impression of neat growth of the discipline, logic, and objectivity, leading to students' adoption of a positivist epistemology, which has been shown to affect their conceptual development. Moreover, the concern with subject matter and the acquisition of what is portrayed as objective knowledge and facts appear to overlook the uncertainty principle in physics whereby its knowledge corpus consists not of absolute truths, but of theories. Critical thinking, which as a result would appear a paramount skill for a physicist, is hardly contemplated in undergraduate curricula, becoming overshadowed by content knowledge (Jones, 2009a, 2009b; Lattuca & Stark, 1994). Consequently, one could argue that teaching approaches based on strong classification and strong framing, driven by a reliance on apparently uncontested and universal knowledge to be assimilated systematically, fail to reflect the social-constructivist epistemology manifest in expert physicists' ways of working. They also fail to give students a holistic view of physics, to include science in the making in addition to ready-made science, and hinder the development of key attributes such as critical thinking, conceptual development, and the ability to tackle problems from multiple perspectives.

How could these two dimensions be reconciled? Referring to the false dichotomy "constructivism versus content," Redish (1999) argues that it is important for students to learn both the process of science and the content, which can be achieved through an approach he calls "scientific constructivism." This entails designing learning environments that encourage students to construct correct scientific ideas through tightly guided discovery, while at the same time covering the subject matter. While it seems taken for granted that students need to learn about the fundamental physical concepts and laws, the appearance of absolute objectivity could be counterbalanced by bringing science and technology studies in the classroom (Mody & Kaiser, 2008), as well as by introducing students to the history and philosophy of science (Fensham, White, & Gunstone, 1994; Matthews, 1994). Extending the science curriculum to integrate these components, students can become aware of how physicists work, of the struggles involved in elaborating theories, of controversies, the "winners" and "losers" among competing theories, and of the fact that knowledge verification and validation contain, too, a human dimension and occur in a specific laboratory, in a specific place and time. In sum, students would learn that theories can, therefore, be prone to error. Making them aware of these facts is one step toward making space for "the awkward student" (Mody & Kaiser, 2008) and toward developing students' critical thinking and conceptual understanding.

Consistency between the epistemology, sociology, and pedagogy of physics has been noted primarily in practices associated with postgraduate-level teaching, characterized by weak classification and framing. The pronounced research preoccupation in physics is reflected in pedagogic approaches. Student supervision is perceived as research rather than teaching, students are integrated in departmental research efforts and their research

is usually closely related to supervisors' specialism. Critical thinking, an essential skill in research, appears to be cultivated in postgraduate students. The "group-based apprenticeship model," contributing significantly to students' socialization into the discipline (Neumann, 2001), mirrors the high level of teamwork encountered in urban disciplines and the collective process of science making. Pedagogic practices at the postgraduate level, driven by students' induction to research, thus seem a faithful reflection of physicists' working environments. A stronger presence of research in the undergraduate curriculum, already advocated in the science education reform literature, could therefore represent another means of narrowing the gap between disciplinary epistemology, sociology, and pedagogy.

CONCLUDING REMARKS

Two overall recommendations emerge from this paper: first, the explicit contemplation of disciplinary epistemology in teaching as a means of avoiding the dangers of epistemological essentialism and, second, the contemplation of the epistemological dimension in pedagogic research.

Students new to a discipline are unaware of the nature of its knowledge, its structure, and the methods involved in its creation, verification, and justification. In the absence of this epistemological foundation, teaching approaches can give students incomplete or inaccurate impressions about a discipline. The natural sciences could thus appear consensual, impersonal, and value-free and isolated from the social or philosophical factors at play. The social sciences, on the other hand, might appear to students as overly divergent, individual, and subjective. Cole's (1983) findings refute these common misconceptions, highlighting that "in the natural sciences, there is probably less consensus at the frontier than has been assumed and that, in the social sciences, there is probably more consensus at the frontier than has been assumed" (p. 134). It therefore emerges as paramount to include in the teaching of a discipline accounts and insights about its history and evolution, the competing theories and the surviving ones, and how that discipline has arrived at its present corpus of knowledge, so as to give students a holistic understanding of their field of study. It also emerges as paramount to familiarize students, already at the undergraduate level, with the practices of knowledge production and validation common in their discipline's community.

These findings also make a case for the contemplation of the disciplinary epistemological dimension in research on teaching and learning, which has often been generic. Epistemological considerations can bring to light nuances able to enrich our understanding of pedagogic approaches across disciplines. These could also contribute to building bridges and facilitating understanding between different disciplinary areas, especially given the increasing interdisciplinarity in higher education and the need for academics in different areas to find common grounds for practice. Therefore, having analyzed the relationship between disciplinary epistemology, sociology, and pedagogic approaches in physics, the paper could be seen as an attempt to shed light on pedagogic idiosyncrasies and increase transparency for other disciplinary communities.

This paper has offered only a bird's-eye view of the relationship between epistemology, sociology, and pedagogy. Further research could investigate more in-depth specific pedagogic aspects or methods in which the social and epistemic interact in physics education. Moreover, the paper acknowledges epistemology to be but one likely influence on academic practice. Therefore, another path for further research could explore the complexity of the reasons behind disciplinary pedagogic approaches, considering not only their epistemological characteristics but also context-dependent social determinants, departmental cultures, and individual factors.

REFERENCES

Abd-El-Khalick, F. (2011). Examining the sources for our understandings about science: Enduring conflations and critical issues in research on nature of science in science education. International Journal of Science Education, 34(3), 353–374.

Abd-El-Khalick, F., & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science: A critical review of the literature. International Journal of Science Education, 22(7), 665–701.

Allen, D. E., Duch, B. J., & Groh, S. E. (1996). The power of problem-based learning in teaching introductory science courses. New Directions for Teaching and Learning, 1996(68), 43–52.

Barad, K. M. (1995). A feminist approach to teaching quantum physics. In S. Rosser (Ed.), Teaching the majority: Breaking the gender barrier in science, mathematics, and engineering (pp. 43–75). New York: Teachers College Press, Columbia University.

Barad, K. M. (2007). Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning. Durham, NC: Duke University Press.

Barton, A. C., & Yang, K. (2000). The culture of power and science education: Learning from Miguel. Journal of Research in Science Teaching, 37(8), 871–889.

Bartsch, I. (2001). Resident alien. A scientist in women's studies. In M. Mayberry, B. Subramaniam, & L. H. Weasel (Eds.), Feminist science studies: A new generation (pp. 30–34). New York: Routledge.

Becher, T. (1990). Physicists on physics. Studies in Higher Education, 15(1), 3–20.

Becher, T., Henkel, M., & Kogan, M. (1994). Graduate education in Britain. London: Jessica Kingsley.

Becher, T., & Trowler, P. (2001). Academic tribes and territories: Intellectual enquiry and the culture of disciplines. Buckingham, England: Open University Press.

Belo, N. A. H. (2013). Engaging students in the study of physics: An investigation of physics teachers' belief systems about teaching and learning physics. Doctoral dissertation, Leiden University, Leiden, The Netherlands. Retrieved June 6, 2013, from https://openaccess.leidenuniv.nl/bitstream/handle/1887/20703/summary.pdf?sequence=6

Bernhard, J. (2000). Improving engineering physics teaching: Learning from physics education research. Paper presented at the Physics Teaching in Engineering Education (PTEE), 2000.

Bernstein, B. (1971). On the classification and framing of educational knowledge. In M. F. D. Young (Ed.), Knowledge and control. New directions for the sociology of education (pp. 47–69). London: Collier-Macmillan.

Bernstein, B. (2000). Pedagogy, symbolic control and identity: Theory, research, critique. Lanham, MD: Rowman & Littlefield.

Blickenstaff, J. C. (2005). Women and science careers: Leaky pipeline or gender filter? Gender and Education, 17(4), 369–386.

Brannigan, A. (1981). The social basis of scientific discoveries. Cambridge, England: Cambridge University Press.

Braxton, J. M. (1995). Disciplines with an affinity for the improvement of undergraduate education. In N. Hativa & M. Marincovich (Eds.), Disciplinary differences in teaching and learning: Implications for practice (pp. 59–64). San Francisco: Jossey-Bass.

Brickhouse, N. W. (2001). Embodying science: A feminist perspective on learning. Journal of Research in Science Teaching, 38(3), 282–295.

Carlone, H. B. (2004). The cultural production of science in reform-based physics: Girls' access, participation, and resistance. Journal of Research in Science Teaching, 41(4), 392–414.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. Journal of Research in Science Teaching, 44(8), 1187–1218

Carter, L. (2008). Sociocultural influences on science education: Innovation for contemporary times. Science Education, 92(1), 165–181.

Carter, L., & Smith, C. (2003). Revisioning science education from a science studies and futures perspective. Journal of Future Studies, 7(4), 45–54.

Cole, S. (1983). The hierarchy of the sciences? American Journal of Sociology, 89(1), 111–139.

Cole, S. (1992). Making science: Between nature and society. Cambridge, MA: Harvard University Press.

Collins, H. M., & Pinch, T. J. (1993). The golem: What everyone should know about science. Cambridge, England: Cambridge University Press.

Cross, A. (1991). The crisis in physics: Dialectical materialism and quantum theory. Social Studies of Science, 21, 735–759.

Crouch, C. H., Fagen, A. P., Callan, J. P., & Mazur, E. (2004). Classroom demonstrations: Learning tools or entertainment? American Journal of Physics, 72(6), 835–838.

Cunningham, C. M., & Helms, J. V. (1998). Sociology of science as a means to a more authentic, inclusive science education. Journal of Research in Science Teaching, 35(5), 483–499.

Danielsson, A. T. (2009). Doing physics-doing gender: An exploration of physics students' identity constitution in the context of laboratory work. Unpublished doctoral dissertation, Uppsala University, Uppsala, Sweden.

DeHaan, R. L. (2005). The impending revolution in undergraduate science education. Journal of Science Education and Technology, 14 (2), 253–269.

Donald, J. G. (1983). Knowledge structures—Methods for exploring course content. Journal of Higher Education, 54(1), 31–41.

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. Studies in Science Education, 38(1), 39–72.

Ehrlich, R. (2002). How do we know if we are doing a good job in physics teaching? American Journal of Physics, 70, 24.

Engelhardt, P.V., Churukian, A. D. & Rebello, N. S. (2012). 2012 Physics education research conference. AIP Conference Proceedings 1513, Philadelphia, PA, August 1–2, 2012.

Etzkowitz, H., Fuchs, S., Gupta, N., Kemelgor, C., & Ranga, M. (2008). The coming gender revolution in science. In E. J. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), The handbook of science and technology studies (3rd ed., pp. 403–429). Cambridge, MA: MIT Press.

Fensham, P. J., White, R. T., & Gunstone, R. (1994). The content of science: A constructivist approach to its teaching and learning. London: Falmer.

Feynman, R. (1965). The development of the space-time view of quantum electrodynamics. Nobel Lecture, December 11, 1965.

Fine, A. (1996). Science made up: constructivist sociology of scientific knowledge. In P. Galison & D. J. Stump (Eds.), The disunity of science: Boundaries, contexts and power (pp. 231–254). Stanford, CA: Stanford University Press.

Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. Science Education, 92(3), 404–423.

Freire, O. Jr. (2003). A story without an ending: The quantum physics controversy 1950–1970. Science and Education, 12(5–6), 573–586.

Galison, P. (1996). Introduction: The context of disunity. In P. Galison & D. J. Stump (Eds.), The disunity of science: Boundaries, contexts and power (pp. 1–36). Stanford, CA: Stanford University Press.

Galison, P. (1997). Image and logic: A material culture of microphysics. Chicago: University of Chicago Press.

Galison, P., & Stump, D. J. (1996). The disunity of science: Boundaries, contexts and power. Stanford, CA: Stanford University Press.

Gooding, D. (1990). Experiment and the making of meaning: Human agency in scientific observation and experiment. London: Kluwer Academic.

Götschel, H. (2011). The entanglement of gender and physics: Human actors, work place cultures, and knowledge production. Science Studies, 24(1), 66–80.

Hake, R. (2002). Lessons from the physics education reform effort. Conservation Ecology, 5(2), 28.

Hammer, D. (1994). Epistemological beliefs in introductory physics. Cognition and Instruction, 12(2), 151–183.

Harding, S. (1991). The "racial" economy of science: Toward a democratic future. Bloomington: Indiana University Press.

Harding, S. (1998). Multiculturalism, postcolonialism, feminism: Do they require new research epistemologies? The Australian Educational Researcher, 25(1), 37–51.

Harding, S. G. (2008). Sciences from below: Feminisms, postcolonialisms, and modernities. Durham, NC: Duke University Press.

Harding, S. (2009). Postcolonial and feminist philosophies of science and technology: Convergences and dissonances. Postcolonial Studies, 12(4), 401–421.

Hativa, N., & Goodyear, P. (2002). Teacher thinking, beliefs and knowledge in higher education. Dordrecht, The Netherlands: Kluwer Academic.

Hativa, N., & Marincovich, M. (1995). Disciplinary differences in teaching and learning implications for practice. San Francisco: Jossey-Bass.

Hermanowicz, J. C. (2006). What does it take to be successful? Science Technology & Human Values, 31(2), 135–152.

Heron, P. R. L., & Meltzer, D. E. (2005). The future of physics education research: Intellectual challenges and practical concerns. American Journal of Physics, 73(5), 390–394.

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. Review of Educational Research, 67(1), 88–140.

Iqbal, H. M., Azam, S., & Rana, R. A. (2009). Secondary school science teachers' views about the "nature of science." Bulletin of Education and Research, 31(2), 29–44.

Jones, A. (2009a). Generic attributes as espoused theory: The importance of context. Higher Education, 58(2), 175–191.

Jones, A. (2009b). Redisciplining generic attributes: The disciplinary context in focus. Studies in Higher Education, 34(1), 85–100.

Jones, M. G., Howe, A., & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. Science Education, 84(2), 180–192.

Kehm, B. M., & Alesi, B. (2010). The implementation of the Bologna process into physics in Europe: The master level. Mulhouse, France: European Physical Society.

Kehm, B. M., & Eckhardt, A. (2009). The implementation of the Bologna process reforms into physics programmes in Europe. Mulhouse, France: European Physical Society.

Kekäle, J. (1999). "Preferred" patterns of academic leadership in different disciplinary (sub)cultures. Higher Education, 37(3), 217–238.

Koponen, I., & Mäntylä, T. (2006). Generative role of experiments in physics and in teaching physics: A suggestion for epistemological reconstruction. Science & Education, 15(1), 31–54.

Kozoll, R. H., & Osborne, M. D. (2004). Finding meaning in science: Lifeworld, identity, and self. Science Education, 88(2), 157–181.

Knorr-Cetina, K. (1995). Laboratory studies: The cultural approach to the study of science. In S. Jasanoff, G. Markle, J. Petersen, & T. Pinch (Eds.), Handbook of science and technology studies (pp. 140–166). Thousand Oaks, CA: Sage.

Kragh, H. (1998). Social constructivism, the gospel of science, and the teaching of physics. Science & Education, 7, 231–243.

Krause, K.-L. D. (2012). Challenging perspectives on learning and teaching in the disciplines: The academic voice. Studies in Higher Education, iFirst, pp. 1–18.

Kreber, C. (2009). The university and its disciplines: Teaching and learning within and beyond disciplinary boundaries. London: Routledge.

Kuhn, T. S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Latour, B. (2008). A textbook case revisited. Knowledge as a mode of existence. In E. J. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), The handbook of science and technology studies (3rd ed., pp. 83–112). Cambridge, MA: MIT Press.

Latour, B., & Woolgar, S. (1986). Laboratory life: The construction of scientific facts. Princeton, NJ: Princeton University Press.

Lattuca, L. R., & Stark, J. S. (1994). Will disciplinary perspectives impede curricular reform. Journal of Higher Education, 65(4), 401–426.

Leach, J., & Scott, P. (2002). Designing and evaluating science teaching sequences: An approach drawing upon the concept of learning demand and a social constructivist perspective on learning. Studies in Science Education, 38(1), 115–142.

Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. Journal of Research in Science Teaching, 29(4), 331–359.

Lee, H., & Witz, K. G. (2009). Science teachers' inspiration for teaching socio-scientific issues: Disconnection with reform efforts. International Journal of Science Education, 31(7), 931–960.

Lepenies, W. (2006). Three and a dozen years. In D. Grimm (Ed.), 25 Jahre. Wissenschaftskolleg zu Berlin 1981–2006. Berlin: Akademie-Verlag.

Lindblom-Ylänne, S., Trigwell, K., Nevgi, A. & Ashwin, P. (2006). How approaches to teaching are affected by discipline and teaching context. Studies in Higher Education, 31(3), 285–298.

Linder, C. J. (1992). Is teacher-reflected epistemology a source of conceptual difficulty in physics? International Journal of Science Education, 14(1), 111–121.

Lising, L., & Elby, A. (2005). The impact of epistemology on learning: A case study from introductory physics. American Journal of Physics, 73, 372.

Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. American Journal of Physics, 74, 118.

Matthews, M. (1997). Introductory comments on philosophy and constructivism in science education. Science & Education, 6(1-2), 5–14.

Matthews, M. R. (1994). Science teaching: The role of history and philosophy of science. London: Routledge.

May, D. B., & Etkina, E. (2002). College physics students' epistemological self-reflection and its relationship to conceptual learning. American Journal of Physics, 70, 1249.

Mayberry, M., Subramaniam, B., & Weasel, L. H. (2001). Feminist science studies: A new generation. New York: Routledge.

McCullough, L. (2004). Gender, context, and physics assessment. Journal of International Women's Studies, 5(4), 20–30.

McCune, V., & Hounsell, D. (2005). The development of students' ways of thinking and practising in three final-year biology courses. Higher Education, 49(3), 255–289.

McDermott, L. C., & Redish, E. F. (1999). RL-PER1: Resource letter on physics education research. American Journal of Physics, 67(9), 755–767.
McGinn, M. K., & Roth, W.-M. (1999). Preparing students for competent scientific practice: Implications of recent research in science and technology studies. Educational Researcher, 28(3), 14–24.
Miller, P. H., Slawinski Blessing, J., & Schwartz, S. (2006). Gender differences in high school students' views about science. International Journal of Science Education, 28(4), 363–381.
Mody, C. C., & Kaiser, D. (2008). Scientific training and the creation of scientific knowledge. In E. J. Hackett, O. Amsterdamska, M. Lynch, and J. Wajcman (Eds.), The handbook of science and technology studies (3rd ed., pp. 377–385). Cambridge, MA: MIT Press.
Moses, I. (1990). Teaching, research and scholarship in different disciplines. Higher Education, 19(3), 351–375.
National Science Foundation (1996). Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology. Washington, DC: Division of Undergraduate Education.
Nelson, D. J., & Brammer, C. N. (2010). A national analysis of minorities in science and engineering faculties at research universities. Retrieved January 7, 2014, from http://faculty-staff.ou.edu/N/Donna.J.Nelson-1/diversity/Faculty_Tables_FY07/07Report.pdf.
Neumann, R. (2001). Disciplinary differences and university teaching. Studies in Higher Education, 26(2), 135–146.
Neumann, R., Parry, S., & Becher, T. (2002). Teaching and learning in their disciplinary contexts: A conceptual analysis. Studies in Higher Education, 27(4), 405–417.
Osborne, J. (2007). Science education for the twenty first century. Eurasia Journal of Mathematics, Science & Technology Education, 3(3), 173–184.
Paty, M. (1999). Comparative history of modern science and the context of dependency. Science Technology & Society, 4(2), 171–204.
Pickering, A. (1984). Constructing quarks: A sociological history of particle physics. Edinburgh, Scotland: Edinburgh University Press.
Pickering, A. (1995). The mangle of practice: Time, agency, and science. Chicago: University of Chicago Press.
Psillos, D. (2004). An epistemological analysis of the evolution of didactical activities in teaching–learning sequences: The case of fluids. International Journal of Science Education, 26(5), 555–578.
Redish, E. F. (1999). Millikan Lecture 1998: Building a science of teaching physics. American Journal of Physics, 67(7), 562–573.
Redish, E. F. (2003). Teaching physics with the physics suite. Hoboken, NJ: Wiley.
Redish, E. F., & Rigden, J. S. (1996). The changing role of physics departments in modern universities. AIP Conference Proceedings 399, College Park, MD, July 31–August 3, 1996.
Redish, E. F., & Steinberg, R. N. (1999). Teaching physics: Figuring out what works. Physics Today, 52(1), 24–30.
Rosser, S. (2012). Breaking into the lab: Engineering progress for women in science. New York: New York University Press.
Ryder, J., & Leach, J. (1999). University science students' experiences of investigative project work and their images of science. International Journal of Science Education, 21(9), 945–956
Sabella, M. S., & Redish, E. F. (2007). Knowledge organization and activation in physics problem solving. American Journal of Physics, 75(11), 1017–1029.
Schwartz, R., & Lederman, N. (2008). What scientists say: Scientists' views of nature of science and relation to science context. International Journal of Science Education, 30(6), 727–771.
Sensevy, G., Tiberghien, A., Santini, J., Laubé, S., & Griggs, P. (2008). An epistemological approach to modelling: Cases studies and implications for science teaching. Science Education, 92(3), 424–446.
Seth, S. (2009). Putting knowledge in its place: Science, colonialism, and the postcolonial. Postcolonial Studies, 12(4), 373–388.
Sin, C. (2012). Researching research in master's degrees in Europe. European Educational Research Journal, 11(2), 290–301.
Smart, J. C., & Ethington, C. A. (1995). Disciplinary and institutional differences in undergraduate education goals. In N. Hativa & M. Marincovich (Eds.), Disciplinary differences in teaching and learning: Implications for practice (pp. 49–58). San Francisco: Jossey-Bass.
Smeby, J.-C. (1996). Disciplinary differences in university teaching. Studies in Higher Education, 21(1), 69–79.
Smeby, J.-C. (1998). Knowledge production and knowledge transmission. The interaction between research and teaching at universities. Teaching in Higher Education, 3(1), 5–20.
Smeby, J.-C. (2000). Disciplinary differences in Norwegian graduate education. Studies in Higher Education, 25(1), 53–67.
Snively, G., & Corsiglia, J. (2001). Discovering indigenous science: Implications for science education. Science Education, 85(1), 6–34.

Songer, N. B., & Linn, M. C. (1991). How do students' views of science influence knowledge integration? Journal of Research in Science Teaching, 28(9), 761–784.

Stathopoulou, C., & Vosniadou, S. (2007). Exploring the relationship between physics-related epistemological beliefs and physics understanding. Contemporary Educational Psychology, 32(3), 255–281.

Stubb, J., Pyhältö, K., & Lonka, K. (2012). Conceptions of research: The doctoral student experience in three domains. Studies in Higher Education, iFirst, 1–14.

Stump, D.J.(1996). From epistemology and metaphysics to concrete connections. In P. Galison & D. J. Stump (Eds.), The disunity of science: Boundaries, contexts and power (pp. 231–254). Stanford, CA: Stanford University Press.

Subramaniam, B. (2009). Moored metamorphoses: A retrospective essay on feminist science studies. Signs, 34(4), 951–980.

Syh-Jong, J. (2007). A study of students' construction of science knowledge: Talk and writing in a collaborative group. Educational Research, 49(1), 65–81.

Tala, S. (2009). Unified view of science and technology for education: Technoscience and technoscience education. Science & Education, 18(3-4), 275–298.

Taylor, P., Gilmer, P. J., & Tobin, K. G. (2002). Transforming undergraduate science teaching: Social constructivist perspectives. New York: Peter Lang.

Thacker, B. A. (2003). Recent advances in classroom physics. Reports on Progress in Physics, 66(10), 1833–1864.

Tobias, S. (1992). Revitalizing undergraduate science: Why some things work and most don't. Tucson, AZ: Research Corporation.

Trowler, P. (2013). Depicting and researching disciplines: Strong and moderate essentialist approaches. Studies in Higher Education, iFirst, 1–12.

Trowler, P., Saunders, M., & Bamber, V. (2012). Tribes and territories in the 21st century: Rethinking the significance of disciplines in higher education. London: Routledge.

Tsai, C.-C. (2006). Reinterpreting and reconstructing science: Teachers' view changes toward the nature of science by courses of science education. Teaching and Teacher Education, 22(3), 363–375.

Tsai, C.-C. (2007). Teachers' scientific epistemological views: The coherence with instruction and students' views. Science Education, 91(2), 222–243

Tuning Project. (2008). Reference points for the design and delivery of degree programmes in physics. Bilbao, Spain: Universidad de Deusto.

Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: high school science and math coursework and postsecondary degree attainment. Journal of Education for Students Placed at Risk (JESPAR), 12(3), 243–270.

Van Heuvelen, A. (1991). Learning to think like a physicist: A review of research-based instructional strategies. American Journal of Physics, 59, 891–897.

Whiteley, P. (1996). The gender balance of physics textbooks: Caribbean and British books, 1985–91. Physics Education, 31(3), 169–174.

Wieman, C. (2007). Why not try a scientific approach to science education? Change: The Magazine of Higher Learning, 39(5), 9–15.

Wong, S. L., & Hodson, D. (2009). From the horse's mouth: What scientists say about scientific investigation and scientific knowledge. Science Education, 93(1), 109–130.

Wong, S. L., & Hodson, D. (2010). More from the horse's mouth: What scientists say about science as a social practice. International Journal of Science Education, 32(11), 1431–1463.

Ylijoki, O. H. (2000). Disciplinary cultures and the moral order of studying—A case-study of four Finnish university departments. Higher Education, 39(3), 339–362.

Yore, L. D., Hand, B. M., & Florence, M. K. (2004). Scientists' views of science, models of writing, and science writing practices. Journal of Research in Science Teaching, 41(4), 338–369.

Does Science Need a Global Language? English and the Future of Research, by Scott L. Montgomery. University of Chicago Press, Chicago, IL, USA, 2013. xiii + 226 pp. ISBN 978-0226535036.

When first reading the title *Does Science Need a Global Language?* one is naturally curious to find out the answer. A reader's second thought may be that the answer is just too simple and that science obviously already *has* a global language, namely English. Scott L. Montgomery goes beyond merely providing an answer to this title. As he puts it himself, "the job of this book is to provide a first-order determination (as we scientists say) of whether a global tongue is truly a good thing for science and why" (p. 23). Indeed, he delves deeply into a myriad of questions that in turn help answer the central one. Through detailed exploration of historical events, geopolitical issues, and personal experiences—and in his own words through "evidence, analysis, and judgment" (p. 168)—he tells the story of how English has become the true lingua franca of modern science and how it compares to other lingua francas of the past. He tells the story of how this phenomenon is affected by and has affected economic, political, educational, and scientific aspects of the developed and developing worlds. He discusses what we can learn from history and what we can predict about the future of a global language and describes the collateral damage a global language can cause in its path of expansion. Finally, Montgomery closes with a resolute answer to his question, ultimately concluding that "participation in global scientific activity means using English" (p. 168).

Regardless of whether the reader shares the author's perspective and agrees with the ultimate conclusion, every chapter presents ample material to sculpt an opinion of one's own. With a topic that could read dangerously dry, Montgomery manages to keep his readers engaged with vignettes about Ben, a scientist from Nigeria; Andre Geim, a 2010 physics Nobel laureate born to German-Russian parents; Roger, an Aboriginal man; and Aziz, a geologist from Egypt. These men, regardless of their education or professional prestige, all agree that speaking English is crucial to succeed in today's global economy, education, or research. Roger, who is not formally educated but whose two sons speak English in addition to three Aboriginal tongues, exclaims that "One never enough!" (p. 103). Ben recognizes that "I would not be hired for these jobs unless I spoke it [English]" (p. 4). On a philosophical note, Andre Geim, whose publications are largely in English, considers a scientist "a worker for all humanity" (p. 70). This mindset seems to be adopted early in life, as an eighth grader from Ethiopia finds that "English is the language of the world, and I want to know the world" (p. 25). The consensus that has apparently reached all corners of the world is that English has become a global language, which renders it "at once denationalized and supranational" (p. 75), no longer requiring anglophone countries as participants for its prestige and its spread, in science or otherwise.

The scientific and industrial revolutions, the building of the British Empire, two World Wars, the end of communism in the Eastern Bloc, and the creation of the European Union are some of the key events that helped propel English into its status as "dominant language of international science" (p. 74). How global is English, then? While it may seem that there is competition from other language "rivals" such as Mandarin Chinese, "educated estimates" (p. 27) indicate that there is in fact no competition at all. In 2010, a quarter of the world's population used English with more than rudimentary skill, and English was declared the official tongue in at least 75 countries that extend beyond the British colonial empire, with astonishing projected numbers of English learners by 2020.

As Montgomery takes the reader through historical events, the influence of English as the dominant tongue on global science and research crystallizes. Increased research and development funding in many developing nations striving for scientific success has led to growing numbers of scientists who use English to communicate across continents and hemispheres. The possibility of communication has allowed for collaborations, overall international participation in science, and multinational research to be "routine, common, or becoming so" (p. 88). This trend is likely to continue, as the desire and possibility of entering careers in science has also been growing globally, demonstrated by the ever-rising numbers of international and graduate students across the world. If we consider workers in the sciences "merchants of knowledge" (p. 75), their trade routes have now expanded beyond the realms of a "dozen or so wealthy nations" (p. 76)—an indication of increased mobility that is proportionate to the rapid growth in scientific endeavors around the globe. English as a lingua franca may serve as an avenue to mobilize thinkers from different cultural and geographic sites and aid in preserving native scientific knowledge.

The book does not present a one-sided analysis, however. Similar to former lingua francas like Greek, Latin, or Arabic, the global status of English as the modern lingua franca of international science may have come at a price. Montgomery cautions that the establishment of any new lingua franca requires "adoption, adaptation, and accommodation" (p. 104), which can give birth to issues of fairness, marginalization, and pro-English bias. Despite the benefits of English as a global language, those who are not native English speakers can be at a strong disadvantage, as developing nations "have less capability to teach and learn English" (p. 105), leading to a situation of haves and have-nots. Those whose native tongue is English, in turn, enjoy an automatic advantage. For science researchers, the difficulty of speaking at academic conferences or writing academic papers in English can be aggravated by large publication databases including English-only journals—a system detrimental to the global visibility and professional progress of scholars who do not publish their research in English but in their native language. As another limitation of its surge as the modern lingua franca, Montgomery respectfully considers English being responsible for the endangerment of indigenous languages and "linguicide" but ultimately dismisses these arguments.

In summary, any lingua franca can bring on "constructive and destructive effects" (p. 158), which are comprehensively and systematically discussed in this book. Do we now know whether science needs a global language? For Montgomery, the benefits clearly outweigh the drawbacks, and his conclusion is clear: "*Yes, it does*" (p. 175). His conviction that science "begs for and even demands such as language" (p. 175) may resonate with the reader based on a synopsis of compelling arguments: English is required for future progress that depends on international collaboration and plurality of participation and is needed so that all nations can participate in and benefit from the science, technology, engineering, and mathematics enterprise.

In this interesting, entertaining, and highly informative read, Scott L. Montgomery teases apart various expected and several unanticipated considerations in determining whether science needs a global language. While some may identify problems with the partly estimated

statistics, others may view the book as food for thought that provides insights into its title question. It is without a doubt a meaningful read for scientists, science educators and researchers, and particularly those interested in science within the context of language and history, as it provides background to why "English is humanity's true global language" (p. 168).

ALEXANDRA O. SANTAU
Department of Instruction and Leadership in Education
Duquesne University
Pittsburgh, PA 15282, USA

DOI 10.1002/sce.21089
Published online 8 November 2013 in Wiley Online Library (wileyonlinelibrary.com).

Science Education and Citizenship: Fairs, Clubs, and Talent Searches for American Youth, 1918–1958, by Sevan G. Terzian. Palgrave Macmillan, New York, NY, USA, 2013. xiv + 235 pp. ISBN 978-1-137-03186-0.

The very first nationwide competition in science for American schoolchildren—the National Science Fair—was held in Philadelphia in 1950. Placed inside the imposing building of the Franklin Institute, finalists from different parts of the country had put their projects on display for the public and members of a jury. It marked the culmination of a phenomenon that started on a small scale but had risen in popularity during the first half of the century. Soon the fair would spread across international borders. Today it constitutes an arena, in the form of the International Science and Engineering Fair, where schoolchildren (and nations) from all over the world compete in science and technology.

The popularity of science fairs in the twentieth century not only marks the success of this idea but also indicates the expansion and changing roles of science education as a societal phenomenon. In the past few decades, researchers from historical, educational, and sociological disciplines have contributed to the considerable growth of historical studies on science education. The cross-disciplinary character of this academic subfield has given multiple insights into the shifting functions and purposes of teaching and training individuals, groups, and nations in science.

Scholarly attention, however, has rarely been directed at events such as those that took place inside the Franklin Institute more than 60 years ago. Fairs, clubs, and competitions existed outside the formal curriculum, but were often initiated and developed in close association with schools and science educators. Sevan G. Terzian's book, *Science Education and Citizenship: Fairs, Clubs, and Talent Searches for American Youth, 1918–1958*, is therefore a welcome and pioneering study with the outspoken aim of delineating the origins and changing profile of extracurricular activities in school science. Terzian's work is important for many reasons: It helps to broaden our understanding of school science and its changing context during the twentieth century as well as explains the institutional expansion of activities outside curricular frames.

The study focuses mainly on the origins and gradually shifting purposes behind fairs, clubs, and talent searches. Science clubs developed at the end of the First World War and were initiated by enthusiastic science educators as a way of furthering the democratic functions of science. Apart from preventing juvenile delinquency and children roaming the streets after school, it was assumed that club activities—often arranged as lectures,

experiments, or excursions to museums or nature areas—would help the students' understandings of modern society as well as the ability to participate in it. Morris Meister was a science teacher who in 1918 founded of the first clubs at Speyer Junior High School in New York. He stated that through insights grounded in careful examination, pupils would develop a more active citizenship in a world where science and technology was growing in importance.

A decade later, local science fairs were organized as more and more club members sought to find places to display their projects. Again it was in New York that standards were set. The Children's Fair opened in 1928 in the American Museum of Natural History, a location that conveyed both cultural and scientific authority. The event was an immediate success. Nearly 3000 local schoolchildren demonstrated their projects, a number that could have been higher had there been more room. Science fairs became increasingly popular during the 1930s as did other similar activities. Terzian's characterization of their growth during the period is rich and detailed. The reader is also given insights into the problems of popularizing science at the time. Constant tensions emerged between enthusiasts and organizations with their own visions and ideas on the one hand and sponsors with funds and more commercial demands on the other.

As the years went by the aim of active citizenship became harder to present as the main reason for the programs. After the Second World War, cold war politics gradually seeped into science education. Estimated manpower shortages of researchers and engineers focused national attention on textbooks and teacher training. This was also true for educational activities outside classroom walls. Terzian shows that extracurricular events as a consequence were caught in a collision between democratic and meritocratic ideals. Soon enough the latter would dominate the scene. This is something that appears most clearly in the third case of the book, scientific talent searches. Launched in the midst of the war, 1941–1942, the plan was to locate and encourage talented youth to seek careers in science or engineering. Initiated by the organization Science Service and its leader Watson Davis, the competition selected its winners based on a range of contributions. In 1942, the participants were told, for example, to write an essay on the topic "How Science Can Help Win the War."

While the main theme in Terzian's book is to point to the origins of extracurricular programs and their subsequent changing profile, the study contributes other interesting and important findings as well. One of them relates to the inequalities in access to clubs, fairs, and talent searchers. Despite the democratic ideals of science education at the time, broad participation could be heavily thwarted. Given the fact that most events took place in cities, schoolchildren from rural areas were not given the same opportunities. At fairs, girls were constantly in the minority and competed in exhibits according to patterns that have lived on to our own days—overrepresented in biology (plants and animals, health, and conservation) and underrepresented in physics, chemistry, astronomy, and geology.

And who was discovered as a "scientific talent"? Watson Davis had stated that the talented youth of the nation should be given opportunities no matter the income of their parents. But even if there was a will to countervail social background, other inequalities were sure to appear. The already low representation of African American entrants in the *Scientific Talent Search* dropped even lower when winners were announced—only 3 of 680 winners (0.4%) between 1942 and 1958. Such numbers were not only a result of poor science education but also a consequence of racially influenced expectations on non-White students.

The way *Science Education and Citizenship* is written brings the reader very close to the circumstances surrounding the above-mentioned activities. The advantage of such a style is that it comes with an assurance of meticulous scholarly work and exhaustive examination.

One of the dangers with such an approach, though, is to get caught in the details, which happens in this book from time to time. The description of single events sometimes becomes a little too specific, and the use of statistics at some points is too plentiful. These parts tend to tire rather than engage the reader.

The author does not make use of an explicit theoretical framework with regard to education or science as societal phenomena. This is certainly not an indispensable component, but at some points one cannot stop to wonder if even more could have been said on such a well-chosen topic. It would also have been interesting to see a study such as this—with new territory broken and being so rich in its findings—relate a little more elaborately to other work within history of science education. These few critical remarks are not meant to discourage anyone from reading *Science Education and Citizenship*. I strongly recommend it to all those with an interest in the arguments for teaching science and the historically changing context of education.

DANIEL LÖVHEIM
Department of Education
Stockholm University
106 91 Stockholm, Sweden

DOI 10.1002/sce.21090
Published online 8 November 2013 in Wiley Online Library (wileyonlinelibrary.com).


ENABLE
DISCOVERY